Supplementary appendix

Table of contents:

Supplementary methods

Selection of patients and samples for the training cohort

The training cohort included patients with confirmed grade 1, 2, 3a FL or undetermined grade (but not 3b) from the PRIMA trial, for whom pretreatment fresh-frozen tissue (FFT) tumor biopsy was available at the Plateforme de Ressources Biologiques, Henri Mondor Hospital, Créteil (bio-bank ID number: BB-0033-00021).

The open-label, international, multicentre randomized PRIMA study enrolled a total of 1,135 patients with untreated high tumor burden FL.[1] Patients were eligible for induction if they were older than 18 years and presented with untreated follicular lymphoma (grade 1, 2, or 3a) diagnosed by a lymph-node biopsy (done within 4 months of study registration). Eligibility required at least one criterion of high tumor burden—namely, bulky disease (one lymphoma lesion greater than 7 cm); three separate nodes of 3 cm or more; symptomatic splenic enlargement; organ compression by tumor, pleural, or peritoneal effusion; raised serum concentrations of either lactate dehydrogenase or β2-microglobulin; or the presence of B symptoms. Patients had to have a performance status of 2 or less on the Eastern Cooperative Oncology Group (ECOG) scale and adequate haematological function (unless due to lymphoma).

During the induction phase, patients were treated with one of three protocol-specified standard immunochemotherapy regimens. The three rituximab combinations used in the PRIMA study were CVP (cyclophosphamide 750 mg/m2 on day 1, vincristine 1·4 mg/m2 [capped at 2 mg] on day 1, prednisone 40 mg/m2 on days 1–5, repeated every 3 weeks for 8 cycles), CHOP (cyclophosphamide 750 mg/m2 on day 1, vincristine 1·4 mg/m2 [capped at 2 mg] on day 1, doxorubicin 50 mg/m2 on day 1, prednisone 100 mg on days 1–5, repeated every 3 weeks for 6 cycles), and FCM (fludarabine 25 mg/m2 on days 1–3, cyclophosphamide 200 mg/m2 on days 1–3, mitoxantrone 6 mg/m2 on day 1, repeated every 4 weeks for 6 cycles). Rituximab (375 mg/m2 at each infusion) was administered on day 1 of each chemotherapy course. Two additional rituximab infusions were administered in patients treated with CHOP (every 3 weeks after the last cycle) and FCM (2 weeks after the first and the fourth cycles) to ensure equivalent exposure to the antibody during induction for all patients.

Response to this induction therapy was assessed 2 to 4 weeks after the last treatment course. After this induction phase, patients who obtained a complete response (CR), an unconfirmed CR (CRu) and a partial response (PR) were randomized in a 1:1 ratio to observation or rituximab maintenance (12 infusions of 375 mg/m2 at 8 week intervals).

Patients were evaluated clinically every 8 weeks during the 2-year maintenance phase and by computed tomography scan every 6 months. Patients with bone marrow involvement at diagnosis underwent bone marrow evaluation at the end of the maintenance phase. Thereafter, a clinical evaluation and a CT scan were performed respectively every 3 and 6 months for a period of 3 years.

RNA extraction from FFT and FFPE samples

RNAs were extracted from FFT samples using the miRneasy Mini kit (Qiagen, Venlo, Netherlands) according to the manufacturer's instructions. RNA quality (RIN) was assessed on Agilent 2100 bio-analyzer. For all cases, additional slides were cut and reviewed by an expert pathologist to ensure representativeness of the samples. Each sample was assessed to contain > 50% lymphoma infiltration.

One to three 10-µm sections were cut from the FFPE blocks according to the size of the biopsy. RNA was extracted for all cases using the Siemens Tissue Preparation Solution (Siemens Healthcare Diagnostics, Cologne, Germany) at the Lymphoma Study Association (LYSA) pathology platform (Henri Mondor Hospital, Créteil).

Expression microarray normalization and probeset filtering

Gene-expression profiling was first performed in the PRIMA training cohort. Of the 160 patients with available FFT biopsy, 149 had sufficient RNA quality (RNA Integrity Number RIN>6·5) to be further processed using Affymetrix U133 Plus 2.0 micro-arrays. One sample was excluded that did not meet criteria after standard Affymetrix quality control.

The data were normalized using GC-RMA. Probesets were annotated using the latest (na34) annotation file from the manufacturer.

For genome-wide analyses, in order to keep only one probeset per Entrez Gene ID, we selected for each gene one "consensus" probeset. This was done as follows:

- When only one probeset was associated with the gene, we selected that probeset.
- When two probesets were associated to the gene, we selected the one with highest variance.
- When three or more probesets were associated to the gene, we computed an "eigen gene" reflecting the main variation in expression in the set of probesets as the first principal component, and selected among theses probesets the one with highest correlation to that "eigen gene".

Finally, we filtered out among the selected unique probesets those that most likely represented noise by discarding probesets with both mean expression below 4 and variance below 0·05 across samples.

This led to a set of 14,356 probesets representing as many unique genes and presumably expressed above noise level.

NanoString expression data quality control and normalization

Expression levels from selected genes were quantified with the NanoString technology, Elements approach (NanoString Technologies, Seattle, WA), based on 50 ng of RNA from FFPE samples both in the PRIMA training set and in the validation sets. To conduct the Elements approach, intermediate probes were designed to target the sequence of interest plus an additional generic tag specific to NanoString generic probes (Tagset). Such intermediate probes were produced by IDT (Leuven, Belgium). The probes sequences used for the 23 genes retained in the final predictor are indicated Table S6. These probes were mixed with the Tagset Elements and RNA prior to 16 hour hybridization at 67 °C using thermocycler. All samples were analyzed in a single multiplexed reaction including 6 negative probes and 6 concentrations of positive control probes. Expression of 95 (training set) or 23 (validation sets) genes and four housekeeping genes (*ACTB, GAPDH, MRPS9, PGK1*) was measured. In order to exclude samples with very low RNA content, we set a minimum count threshold of 6,500 for the sum of all four housekeeping counts. We thus excluded 29 samples, out of the 490 samples from all three validation cohorts.

The data were normalized as follows:

- the mean of negative control values were substracted for each sample, the minimal count after this step was set to 1.
- The data were log2 transformed
- The mean of the log2 counts of housekeeping genes was subtracted for each sample (this is equivalent to dividing by the geometric mean of the four housekeeping genes in the non-log scale)
- In order to have normalized values in approximately the same range as the original counts, we added $\log_2(15{,}000)$ to all genes.

The NanoString experiments for the MAYO and BCN cohorts were performed at a later date on a different experiment. To adjust for any technical variation between these two codesets, replicates of a control sample (Universal RNA) were included during each NanoString run. For each gene, we estimated the shift in mean of the normalized levels of the control samples with the new codeset and used these gene level adjustments to correct the normalized levels of the samples assayed with the new codeset.

Model development and validation

PFS-supervised analysis

A PFS-supervised analysis was conducted in the 134 randomized patients of the training cohort, for whom Affymetrix microarray gene expression profiles measured on FFT sample were available. All genes were then tested for association with PFS, using a model including the effect of maintenance treatment as a cofactor and applying Benjamini Hochberg multiple testing adjustment (likelihood ratio test comparing model with maintenance and gene expression to model with only maintenance, 5% False Discovery Rate). We identified a list of 395 genes whose expression levels were associated with PFS.

Filtering of candidate genes and technical replication between Affymetrix and NanoString values

In order to build a predictive model for PFS applicable to FFPET samples with a limited number of genes, we selected a sub-panel of 95 genes that were further profiled using NanoString technology on FFPE biopsies from 53 patients of the training set (47 patients were part of the randomized cohort and 6 were not randomized). Those samples were used to assess the replicability of RNA measurements for the different genes between FFT samples microarrays and FPPE NanoString-based measurements.

Of the 395 genes significantly associated with PFS in supervised analysis, a sub-panel of 95 genes was carefully curated to select a panel 95 genes to be further quantified by NanoString technology (Figure S2).

A first filtering step was performed by comparing the individual prognostic value of each gene (ie, whether a higher expression was associated with a longer or shorter PFS) obtained on microarrays with RNA-sequencing data available for all samples of the training cohort (n=148)[2]. This led us to discard 91 genes out of 395 because their expression quantified by RNA-seq did not confirm the impact on PFS identified from microarrays. This information is shown for each gene in Table S2.

The second step consisted in selecting genes with high variance (interquartile range IQR>0·3), thus excluding 9 genes.

The third filtering step excluded three genes because of discordance between the prognostic values of the different probesets for a given gene (from the microarray).

Finally, 95 genes were selected out of the 292 remaining genes by integrating different parameters:

- genes with a high mean expression across samples (>7·5) were favored
- genes with a high variance (IQR>1) were favored
- the reproducibility between different probesets (several probesets with highly concordant prognostic values were favored)
- the p-value obtained from the PFS-supervised analysis (genes with p-value<0·02 were favored)
- biological relevance (eg, known function in B-cell development, known prognostic value in other cancers…) was taken into account (scored for each gene as “high”, intermediate” or “low”).

This panel of 95 genes was then assayed using NanoString on the 53 FFPE samples and Pearson correlation coefficient was used to compare measurements with those obtained on the 53 matched Affymetrix arrays. Individual gene correlations between both technologies ranged from -0·2 to 1. Genes with correlation coefficient >0·75 were then retained for the model (n=23) (Figure S4A).

Model fitting

We used these 23 genes with good technical replication to fit a penalized multivariate Cox model on the Affymetrix-frozen 134 randomized patients including maintenance treatment as a non-penalized cofactor.

The R package glmnet[3] was used to apply Ridge penalization with variable standardization. The penalization parameter lambda was chosen using 10-fold cross validation and the one-standard-error rule.

This model allowed us to define a multi-gene score predicting the PFS outcome of patients independently of the effect of the maintenance treatment.

Model validation in NanoString data

In order to apply the model with the gene weights and the threshold determined on Affymetrix measurements, we transformed ("recalibrated") the NanoString measurements to set them to the same scale as the corresponding Affymetrix measurements. This was done using the 53 samples from the training cohort that had been assayed with both technologies: for each gene we calculated the shift in mean level and the scaling factor in standard deviation needed to linearly transform the distribution of NanoString measurements of that gene across the 53 samples to match the mean and standard deviation of the Affymetrix measurements on those same samples. These gene-specific shifts in mean and scaling factors were then used to recalibrate all subsequent normalized NanoString measurements in the validation cohorts. Next we computed the score of each patient using the gene weights defined by the penalized model fitted on the training cohort and the recalibrated expression values.

An optimal threshold on the score to separate patients into low- and high-risk groups was determined using the maxstat package (http://CRAN.R-project.org/package=maxstat, R package version 0.7-24)[4] to select the cut-off value producing the maximal logrank score in the training cohort.

We then computed the C-index of this score on our training and validation cohorts, the logrank p-values based on the thresholds defined in the training cohort and the Wald p-value associated with the contribution of the signature score in a Cox model including maintenance treatment and FLIPI-1 as a co-factors.

Unsupervised analysis and functional enrichment analyses

Gene clustering was performed on expression data from Affymetrix microarrays in the training cohort according to an Independent Component Analysis (ICA) using the fastICA algorithm.[5–7] ICA identifies expression signatures ("components") capturing major underlying sources of variation in the data and attributes a weight for each gene in each component, which can be positive or negative. The activity of each component in each sample can be computed as the weighted sum of the expression of its component genes ("metagene") while the genes most influential in each component can be retrieved by selecting those with the largest weights.

fastICA parameters

The fastICA algorithm requires a number of components to be specified as a parameter. In order to determine the optimal number of components that could be extracted from our datasets, we used an approach similar to Rotival et al.[5] Since the first step of fastICA is to transform the data using singular value decomposition (SVD) into the specified number of components, we calculated the variance explained by each SVD component in our data and compared that to the variance explained by SVD components in random (permuted) data. According to this analysis (data not shown), the variance explained by each of the 25 first components in the original data was superior to the variance explained by the corresponding components in permuted data. We therefore determined that the 25th first components did correspond to real signal in the data and we used 25 as the number of ICA components in the fastICA algorithm.

The R package fastICA was used to perform the analysis, using the "parallel" version of the algorithm, the "logcosh" function, with n.comp=25, alpha=1 and tol=1e-04. The data was standardized per gene to mean 0 and variance 1 prior to applying the algorithm.

Stability of the fastICA analysis

Since fastICA is an iterative algorithm that starts with a random initialization state, it is necessary to ensure that the solution of the algorithm for which convergence was reached is stable across multiple runs. We therefore ran fastICA ten times and compared the components output at each run both in terms of their leading genes (S matrix) and in terms of the correlation of their activity across samples (A matrix). We observed that for all runs, very similar components were found and we selected for our analyses one of the ten runs for which all 25 components were repeatedly identified in multiple runs (data not shown).

Leading genes of components and orientation of the components

We defined for each component the set of "leading genes", those most influential in the component, as those with weights over 3 standard deviation in absolute value. In order to facilitate their interpretation, some components were flipped (the sign was changed) so that for every components there were more leading genes with positive weights than leading genes with negative weights. This was made so that the positive score of a

component would always be associated with positive expression of most of its leading genes, thereby facilitating the interpretation.

Experimental - batch associated components

Each ICA component score (A matrix) was tested for association with experimental factors: date of RNA extraction and date of hybridation of the arrays using ANOVA. For 6 of the components, R2 quantifying the variance explained by one of these two factors was over 25% and those were therefore discarded for being associated with batch effects and reflecting experimental rather than biological variations (ICA1, ICA2, ICA3, ICA11, ICA14, ICA21).

Functional enrichment analyses

To infer potential biological relevance of ICA components, we compared their leading genes with the Molecular Signature Database (MSigDB) v5.0 and LLMPP collections of precompiled gene sets (SignatureDB).[8] Significant overlaps between components and gene sets were evaluated using one-sided Fisher Test and applying multiple-testing correction to account for the number of gene sets within each collection.[10]

Calculation of IR1 and IR2 signatures scores

In order to reproduce as faithfully as possible the IR1 and IR2 signatures described by the LLMPP group,[11] we extracted the Affymetrix probesets (originally from Affymetrix chips U133A and U133B but also present on our Affymetrix U133plus2 chips). We computed the mean GC-RMA value for the two signatures for all samples across all genes in both signatures.

Clinical correlations and survival analyses

Patient baseline characteristics were compared between groups using the Fisher's exact test or Student t test when appropriate. Time-to-event parameters were estimated by the Kaplan-Meier product limit method and compared by log-rank test, or using a multivariate Cox model when mentioned. Regarding the training cohort and the validation cohort from the PRIMA trial, overall survival (OS) and progression-free survival (PFS) could be evaluated from the date of registration before induction phase or from the time of randomization to rituximab maintenance or observation.

We evaluated the association between early progression of disease (POD) and OS as previously described.[12] Two groups were defined: patients with early POD 2 years or less after diagnosis (n=103) and those without POD within 2 years, the reference group (n=335). Patients lost at follow-up or died less than 2 years after diagnosis without progression were excluded (n=22). The risk-defining event was survival from time of POD for early progressors or from 2 years after diagnosis for the reference group.

For determining the ability of our score to predict POD24, sensitivity, specificity, positive and negative predictive values (PPV, NPV) were calculated as follows: Sensitivity=TP/(TP+FN), Specificity=TN/(FP+TN), PPV= TP/(TP+FP), and NPV= TN/(TN+FN), where TP represents the number of patients predicted high risk with our score that are early progressors, FP the number of patients predicted high risk that are not early progressors, TN the number of patients predicted low risk that are not early progressors, and FN the number of patients predicted low risk that are early progressors.

All survival curves were plotted using the survminer R package version 0.3.1 (https://cran.r-project.org/web/packages/survminer/index.html).

Evaluation of ICA13 signatures in public datasets

The ICA13 signature was assessed in the following GEO datasets: GSE2350,[13] GSE14714,[14] GSE56314.[15]

GSE2350 pre-processing

For this series, two pre-processed expression tables were downloaded from GEO, corresponding to samples analyzed with two different versions of the Affymetrix array (HGU95A and HGU95Av2). These two datasets were combined using all common probesets between the two microarray versions, and a quantile normalization[16] was performed to homogenize the two series followed by a log2 transformation. Next, using the latest annotations from Affymetrix (HG_U95Av2.na36.annot), for each leading gene from the ICA13 component we selected the corresponding probeset that displayed the largest variance. In total we were able to match 59 out of 133 leading genes with positive weights, and 11 out of 21 leading genes with negative weights.

GSE14714 pre-processing

The pre-processed data from this series was obtained from GEO. Next, using the annotation file from the EBI website (http://www.ebi.ac.uk/arrayexpress/arrays/A-GEOD-3278/?ref=E-GEOD-10793) corresponding to this microarray (NCI/Staudt human 15K v13), we were able to retrieve expression for 49 of the ICA13 leading genes with positive weights and 9 of those with negative weights. When several probesets matched the same gene, the probeset with largest variance was selected.

GSE56314 pre-processing

The pre-processed data from this series was downloaded from GEO. Since this series was analyzed using the same Affymetrix microarray as our training cohort (Affy U133plus2), we were able to directly extract the exact same probesets for all leading genes of the ICA13 component.

Computation of an average ICA score

In all three datasets, the matched genes were combined into a summarized score by taking the average standardized expression, weighted by (+1) for genes with a positive weight, and by (-1) for genes with a negative weight.

Supplementary results

Patients' characteristics in the training cohort

There were no significant differences in the patient clinical characteristics between the training cohort and the remainder of the patients in the PRIMA trial (Table S1), except for the more frequent presence of B symptoms in patients of the entire PRIMA trial population (32% vs 22%; P=0·04). We evaluated the primary end point (PFS in randomly assigned patients) of the PRIMA study in this subpopulation. Patients in the rituximab maintenance group had significantly longer time to progression than patients in the observation group (P=0·0088, Figure S1), thereby in line with the results of the whole PRIMA study.[1]

Analysis of IR1 and IR2 signatures

When combining both signatures using the exact same weights as in the Dave publication[11] (-2·36* IR1 + 2·71*IR2), this composite score was not associated with PFS (P=0·77). We then used a multivariate Cox model with both IR1 and IR2 as continuous variables. In this model, the coefficients attributed to the signatures were -1·635*IR1+0·363*IR2 but only IR1 remained associated independently with PFS (P=0·0062) while IR2 did not retain significance (P=0·43).

Next, both signatures were evaluated individually by separating samples using a median threshold and performing a log-rank test (Figure S12). For both signatures, the higher expression population had a significantly more favorable PFS.

Supplementary Figures

Figure S1: Kaplan-Meier estimates of progression-free survival (PFS) according to treatment group in the training cohort.

We evaluated the primary end point (PFS in randomly assigned patients) of the PRIMA study in the group of randomized patients included in the PFS-supervised analysis (n=134). As in the original PRIMA cohort, patients treated with maintenance by rituximab had a significantly different PFS from those receiving no maintenance (P=0·0088, log-rank test). For each time point are indicated the number of patients at risk and (number of patients censored).

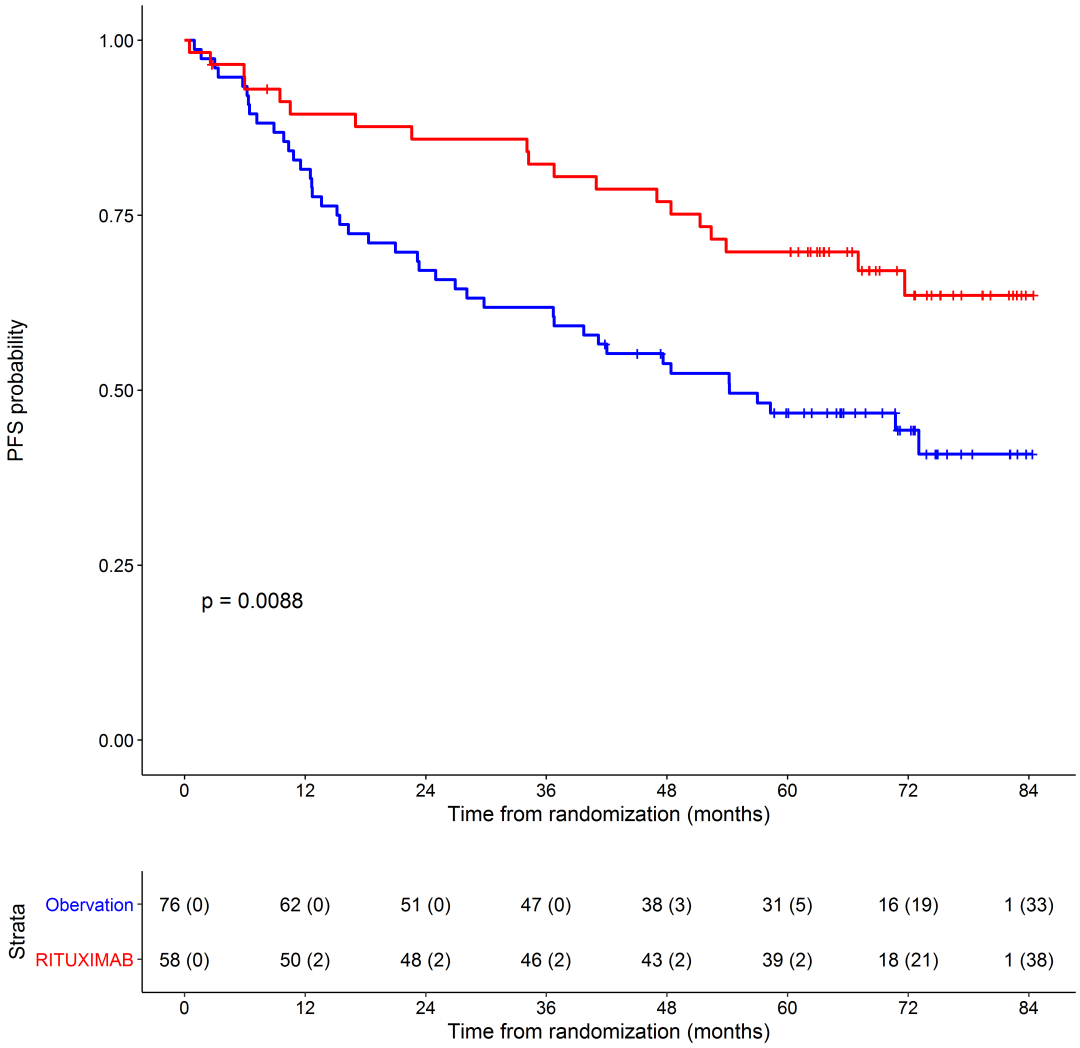


Figure S2: **Selection criteria of the 95-gene subpanel for quantification by NanoString.**

PFS-supervised analysis identified 395 genes whose expression predicted progression. This was compared to our data obtained by RNA sequencing and only genes with concordant prognostic value were selected. The gene list was further refined by eliminating genes with low variance (IQR<0·3) and genes with multiple, discordant probesets. Finally, the last step of selection consisted in manual curation, taking into account technical and biological criteria. See supplemental methods for details. *Abbreviations: IQR, inter-quartile range.*

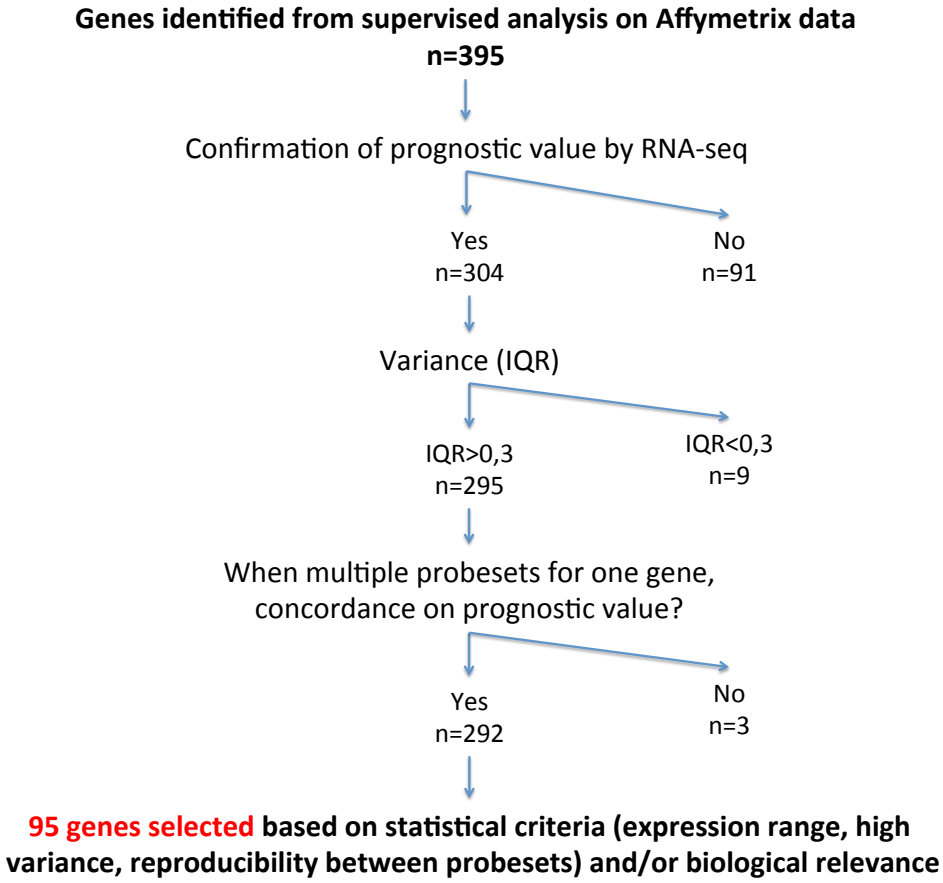


Figure S3: Predictive model building in the training cohort.

Hierarchical clustering of randomized patients from the PRIMA training cohort. A PFS-supervised analysis identified 395 differentially expressed genes that formed a globally unidirectional signature with all genes falling into two tight and anti-correlated clusters. Higher expression levels were associated with longer PFS for 228 genes ("Good prognosis"), and with shorter PFS for 167 genes ("Bad prognosis"). The Euclidean distance metric was used for columns (samples), while the Pearson metric was used for rows (genes). Ward agglomeration was used for building trees.

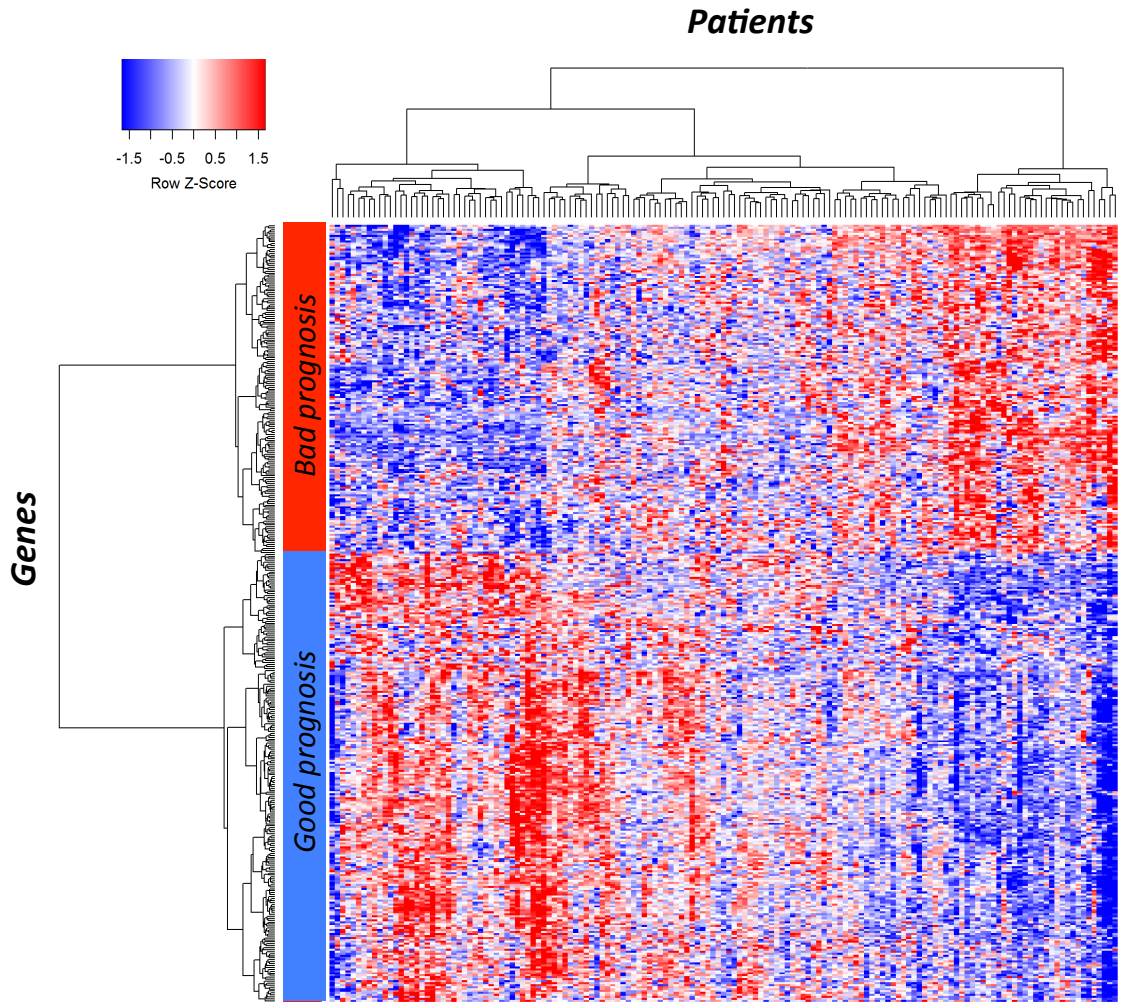


Figure S4: Technical correlation between Affymetrix and Nanostring measurement.

A: Correlation values for the 95 genes quantified using both Affymetrix microarray (performed on RNAs from fresh-frozen tissues) and Nanostring technique (performed on RNAs from paraffin-embedded tissues). Only genes with correlation coefficients >0·75 (red dotted line) were retained in the final predictor.

B: Correlation of the 23-gene score between Affymetrix microarray and NanoString technology. Technical replication between the two technologies was performed for 53 patients from the training cohort. The optimal threshold of 1·075 that separates patients into high- and low-risk groups is represented as a dotted red line.

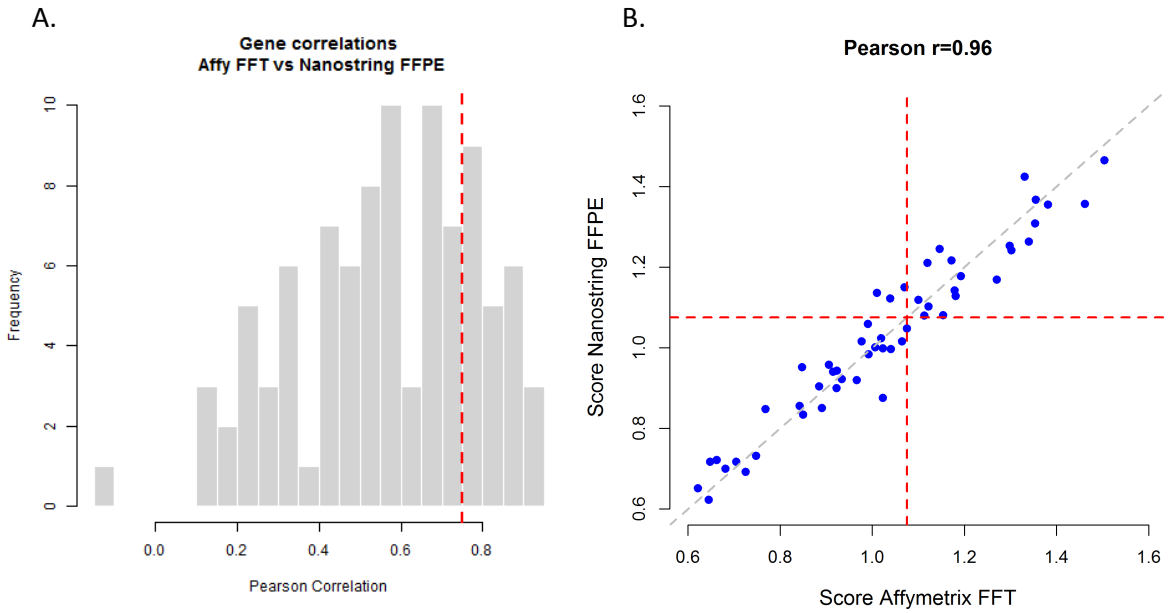


Figure S5: Determination of a threshold value discriminating two prognostic groups in the training cohort.

A: An optimal threshold was determined using the MaxStat package, which selected the cut-off value that produced the maximal log-rank score.
B: Distribution of the individual patient scores in the training cohort. The vertical dotted red line is placed at the optimal cut-off that separates patients into low- (n=87) and high-risk (n=47) groups.

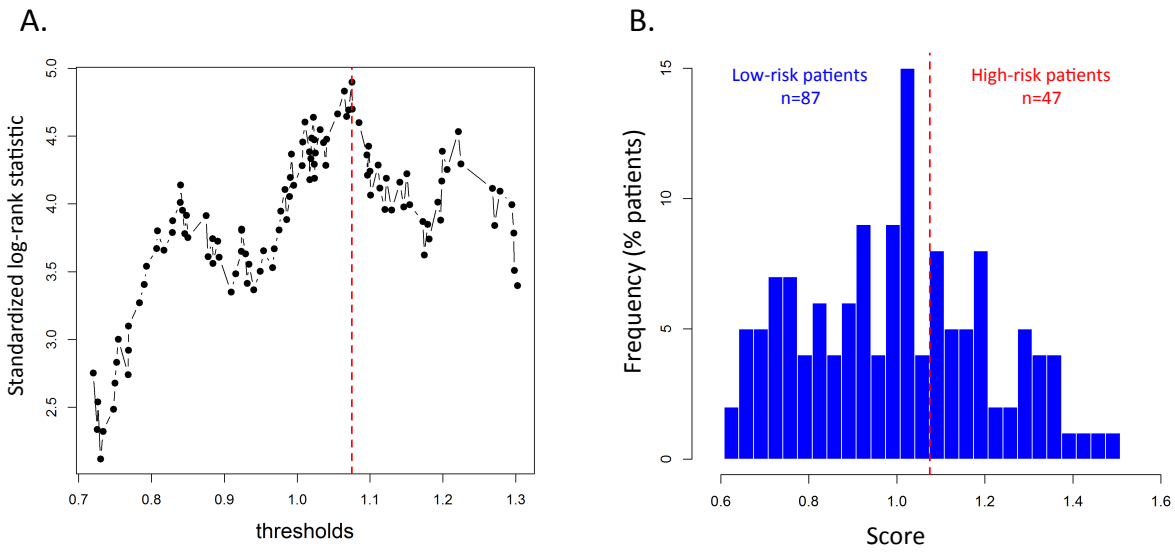


Figure S6: Comparison of Area Under the Curve (AUC) estimates in the combined validation cohort for both FLIPI and the 23-genes score prediction of PFS.

AUC are shown at 0.1 years spaced time-points between 0 and 10 years. These estimates were obtained using the Nearest Neighbour Estimation method[17] as implemented in the survival ROC R package.

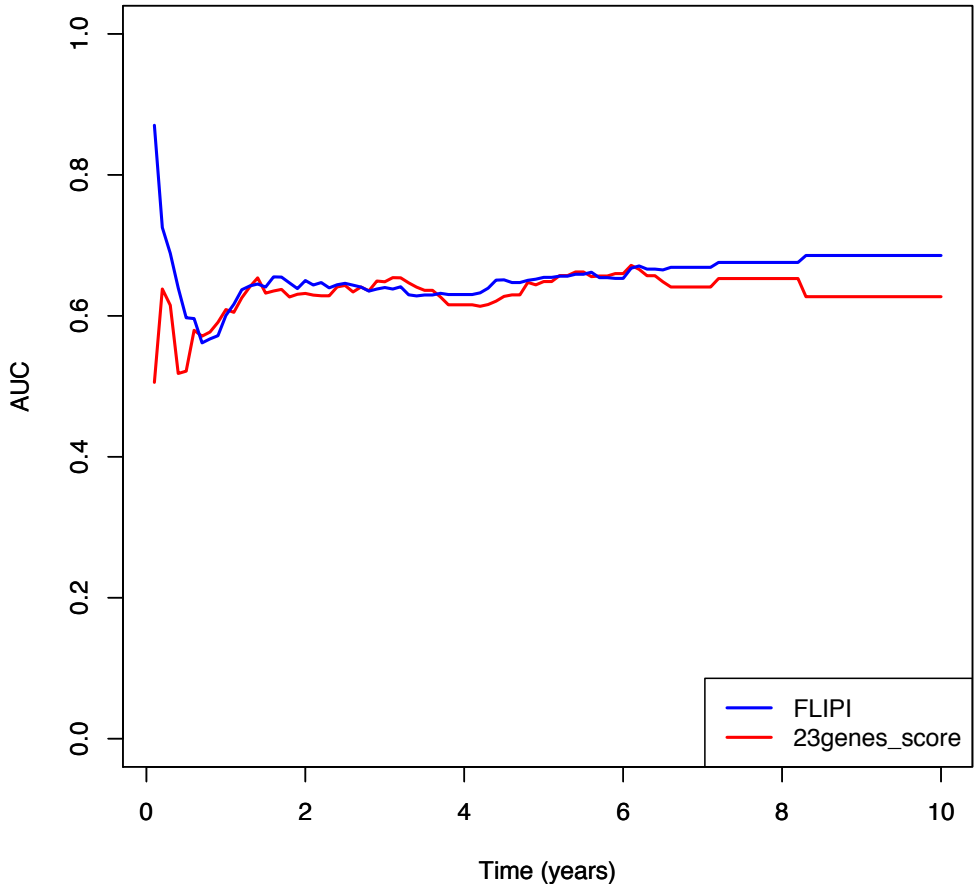


Figure S7: **Kaplan-Meier estimates of progression-free survival (PFS) according to the predictor score in each induction treatment group.**

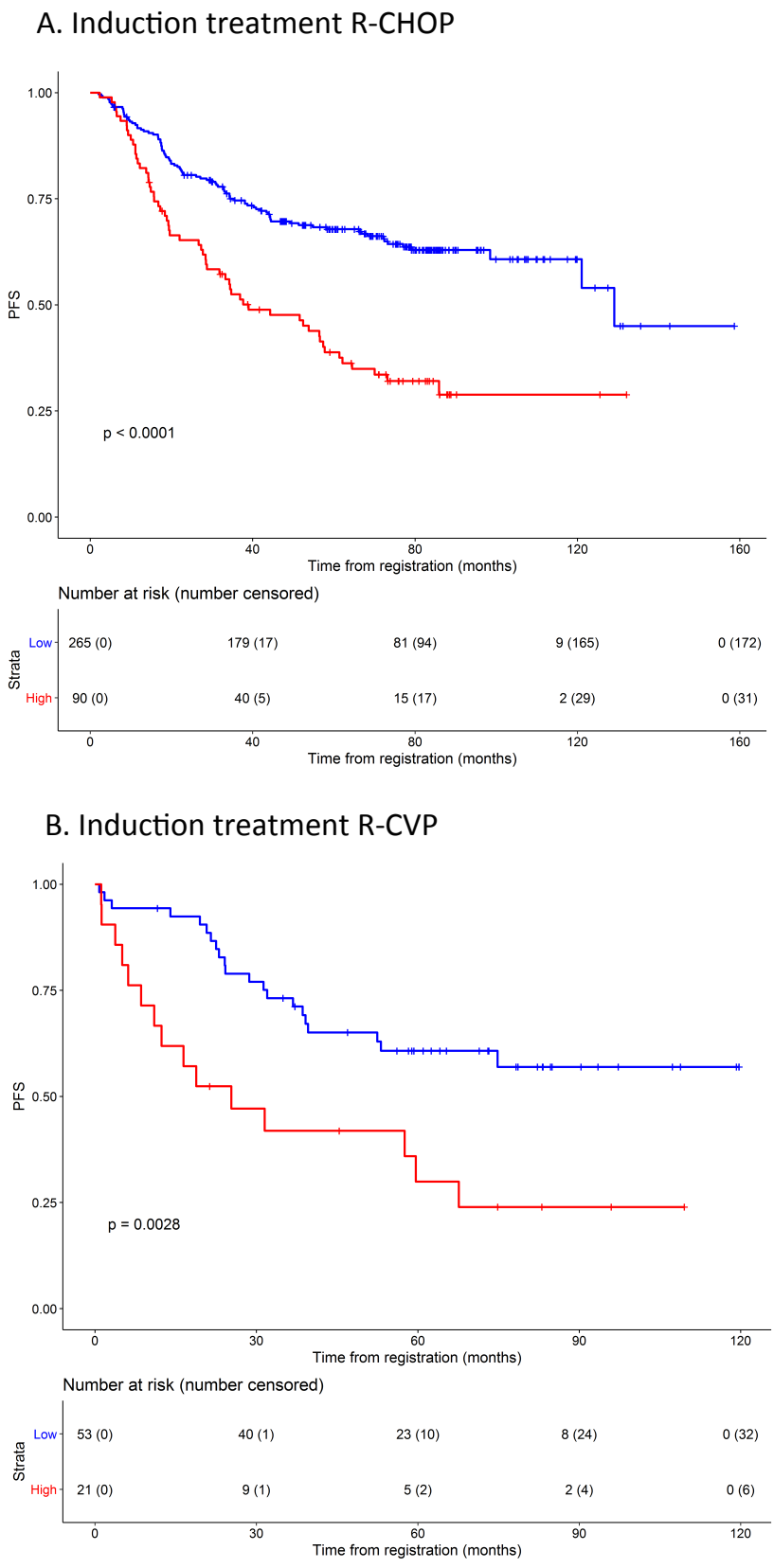


Figure S8: **Kaplan-Meier estimates of overall survival according to POD24 status in the overall validation cohort.**

The POD24 status is defined either at the time of progression, if this happens within 2 years after diagnosis (early progression), or at the 2-year time point if no progression occurred (see supplemental methods). Events (deaths) are then recorded starting from this time point (date of progression if happened before 2 years, or starting 2 years after diagnosis). Patients with early progression (early progressors, black curve) had a significantly shorter time to overall survival than patients without progression within 2 years (reference group, grey curve) (P<0·0001, log-rank test).

Abbreviations: POD24: Progression Of Disease at 24 months.

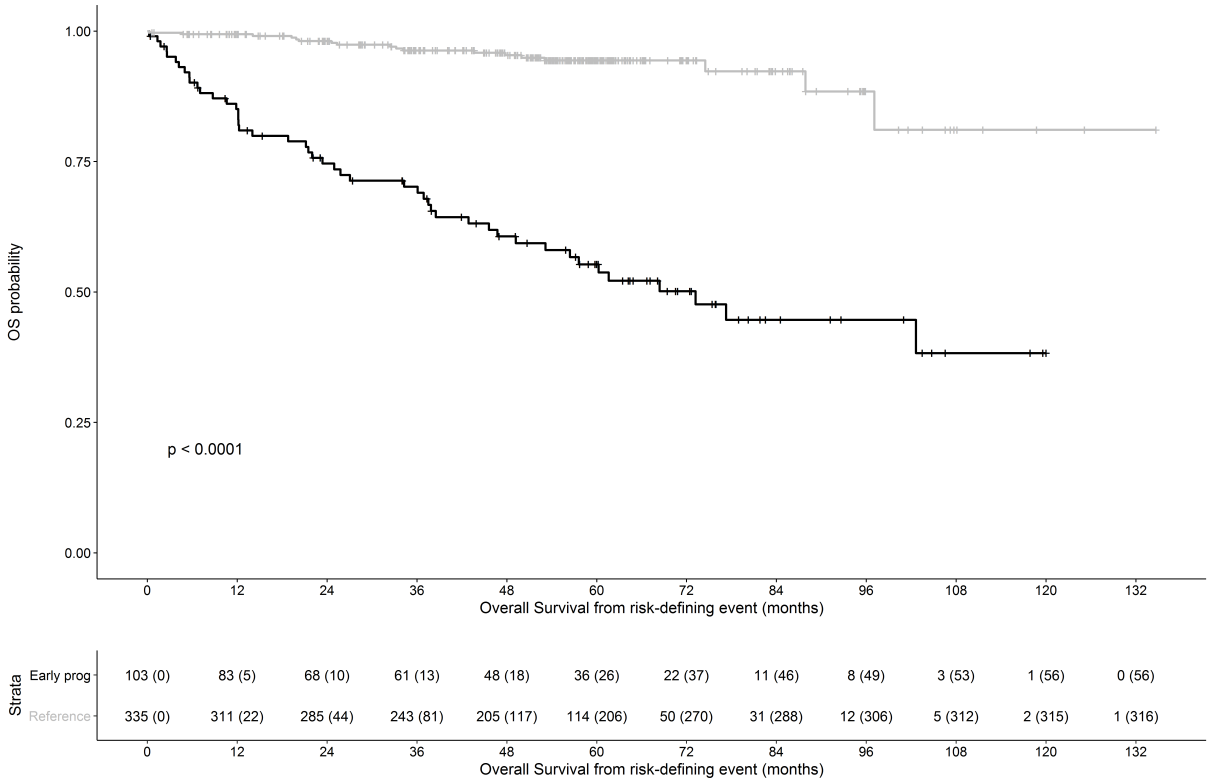


Figure S9: **Kaplan-Meier estimates of overall survival (OS) according to signature score in the training and the combined validation cohorts.**

The 23-gene score did not predict the risk of death, neither in the training cohort (A) nor in the validation cohorts (B). The red and blue curves correspond to patients at high and low-risk, respectively. In contrast with Figure S8, the starting point for the survival curves is the date of diagnosis.

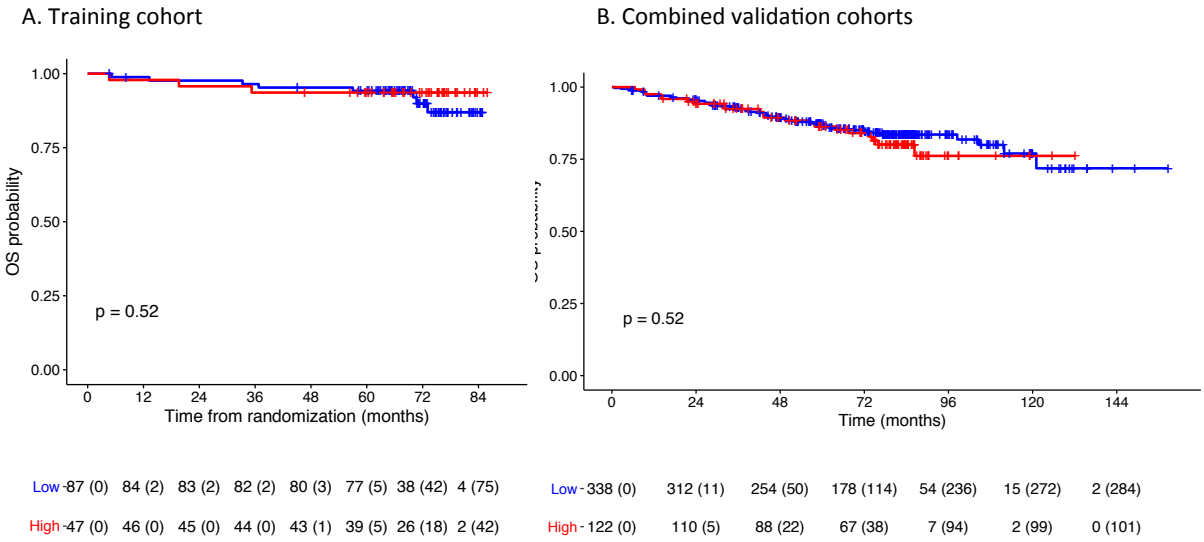


Figure S10: **Kaplan-Meier estimates of lymphoma specific survival according to signature score in the combined validation cohorts.**

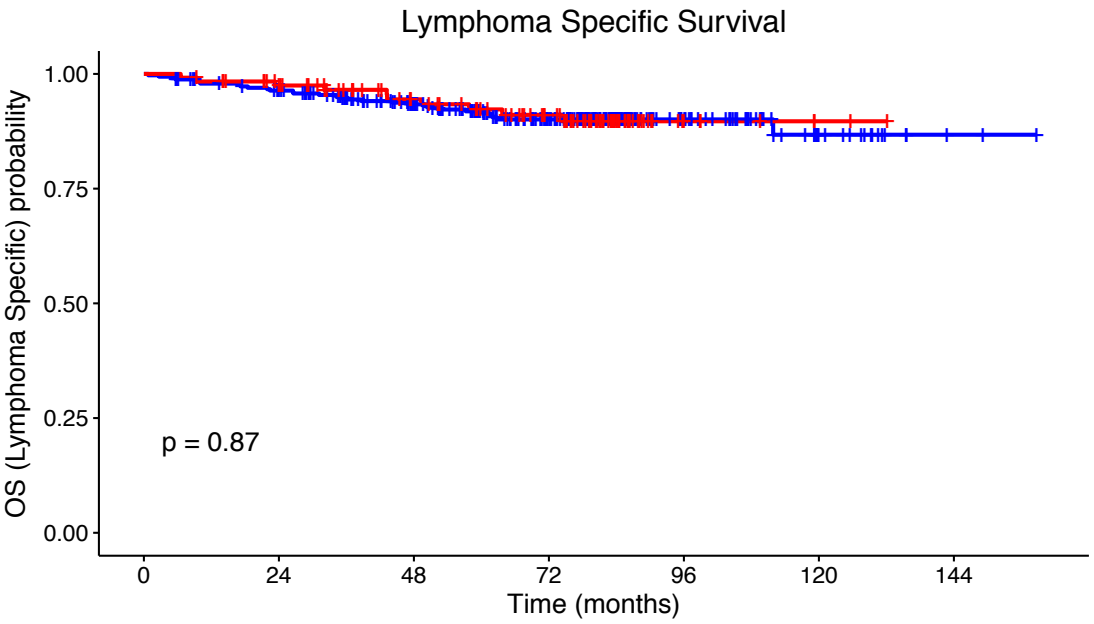


Figure S11: **Cellular origin of the expression predictor genes.**
This heatmap shows the expression of the 23 genes from the predictor in 5 different types of samples: CD19+ sorted B cells from FL tumours (FL_B cells, n=23), CD19+ sorted B cells from reactive lymph nodes (RLN_B cells, n=10), CD19-CD22- sorted cells from FL tumours (FL_Env, n=15), CD19-CD22- sorted cells from reactive lymph nodes (RLN_Env, n=6), and CD19-CD22- sorted cells from reactive tonsils (Tons_Env, n=5). The data were obtained from Pangault et al.[18] (GSE107367) and were GC-RMA normalized and scaled (mean 0, variance 1). *Abbreviations: RLN, reactive lymph node ; FL, follicular lymphoma ; Tons, tonsils.*

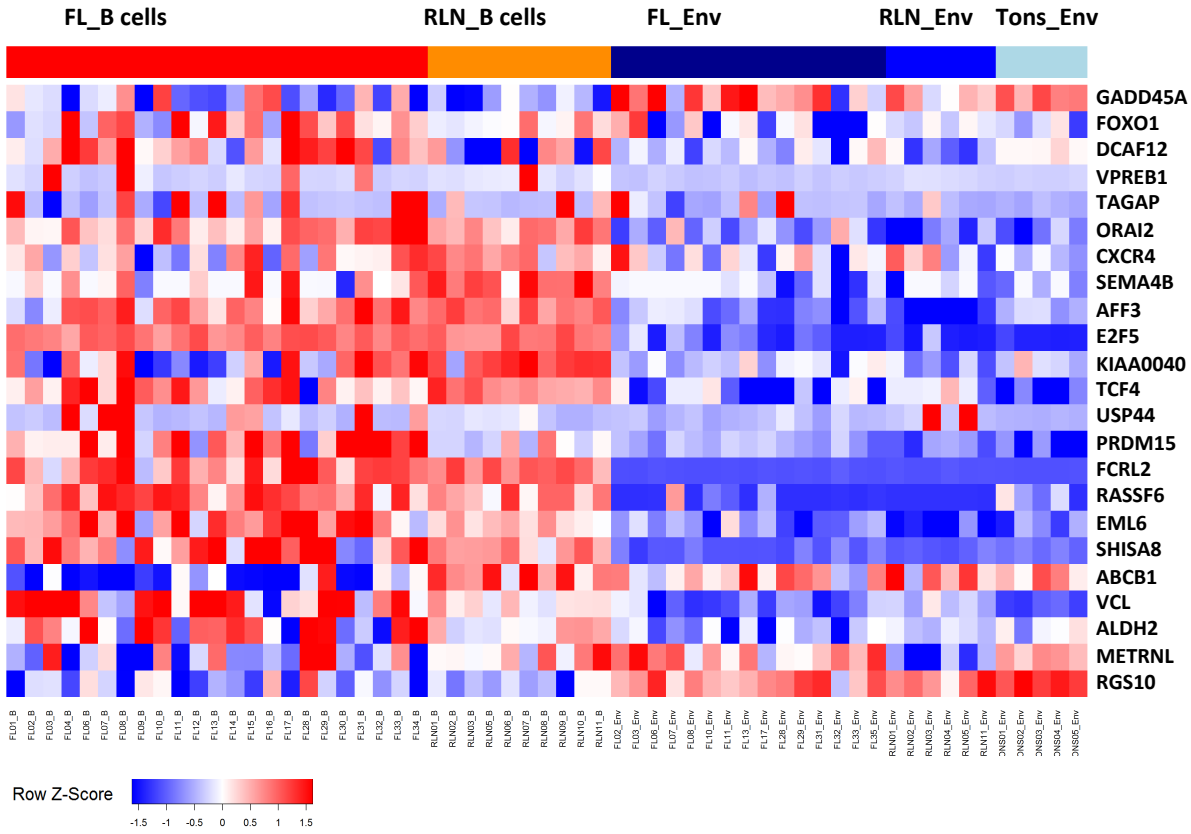


Figure S12: **Kaplan-Meier estimates of progression-free survival according to the signatures reported by the LLMPP group**.

The IR1 and IR2 signatures were evaluated for association with PFS among randomized patients of the PRIMA training cohort. When combining both signatures using the exact same weights as described by the LLMPP group[11] (-2·36*IR1 + 2·71*IR2), this composite score was not associated with PFS. Both signatures were then evaluated individually by separating samples using a median threshold and performing a log-rank test (red curve: high score, blue curve: low score). In our cohort, both signatures were associated with a significantly more favorable PFS (IR1, p=0·0056; IR2, p=0·0018). *Abbreviations: PFS : Progression-free survival.*

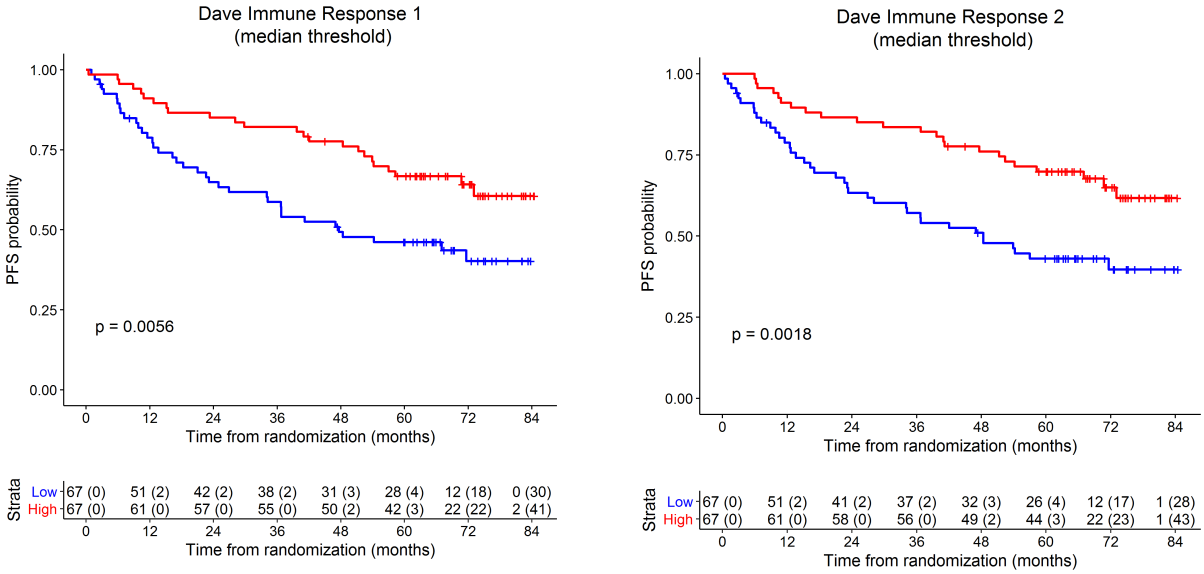


Figure S13: Expression signatures from FL cells and their microenvironment.

A. Hierarchical clustering of 148 FL patients from the training cohort according to the score of expression components identified by unsupervised analysis. In a sample (column), each component (row) is represented by the weighted expression of its leading genes. Clustering was performed using the Pearson correlation for rows, the Euclidean distance for columns and Ward agglomeration. B. Kaplan-Meier estimates for progression-free survival according to the score of the ICA13 component. Patients with an ICA13 score above the median (red curve) had a significantly shorter time to progression than patients with an ICA13 score below the median (blue curve) (p=0·00091, log-rank test). C. Correlation between the 23-gene predictor and the ICA13 signature. The two scores were highly correlated (r=0·79, p<0·0001, Pearson correlation test), suggesting a common biological basis. *Abbreviations: ICA: independent component analysis, PFS: progression-free survival.*

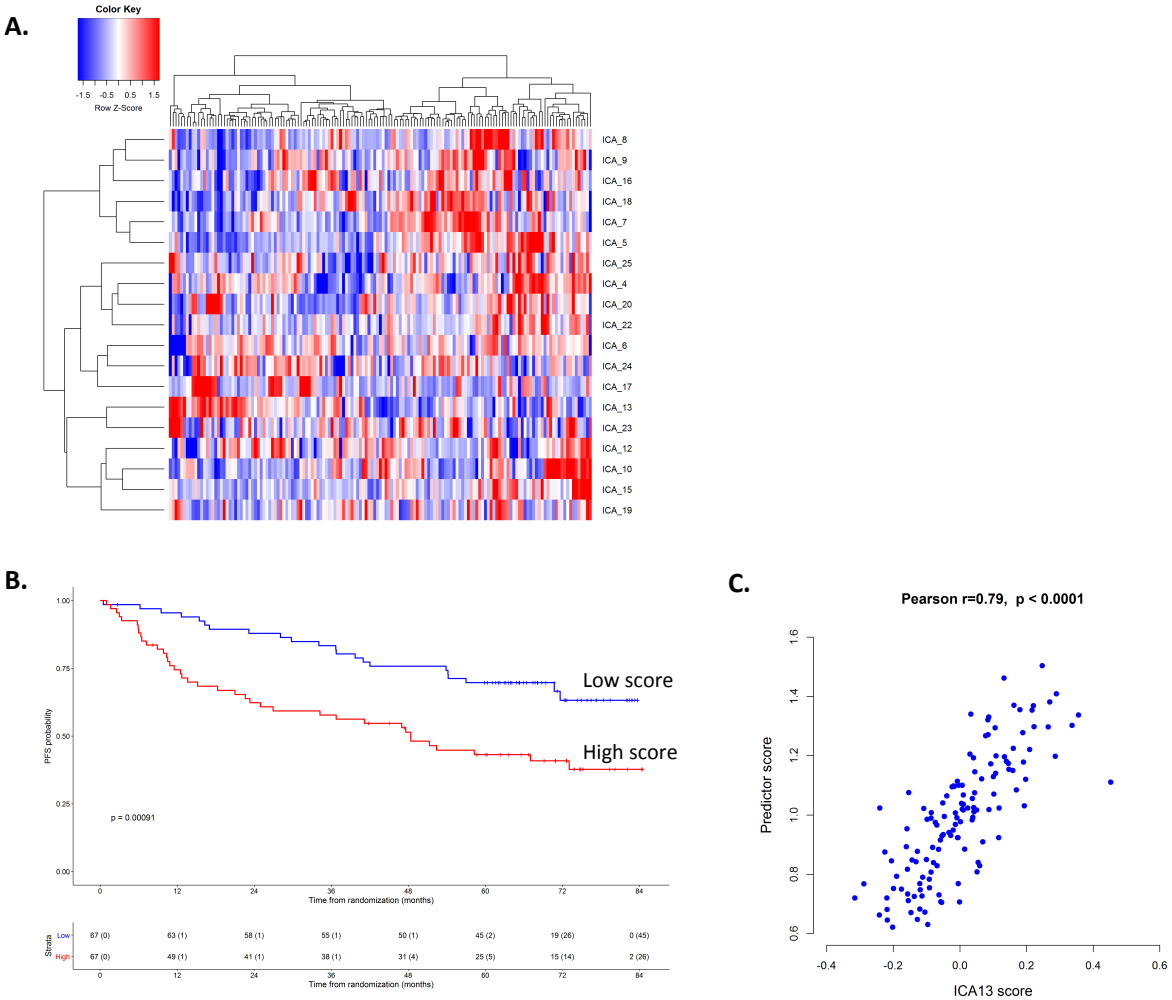
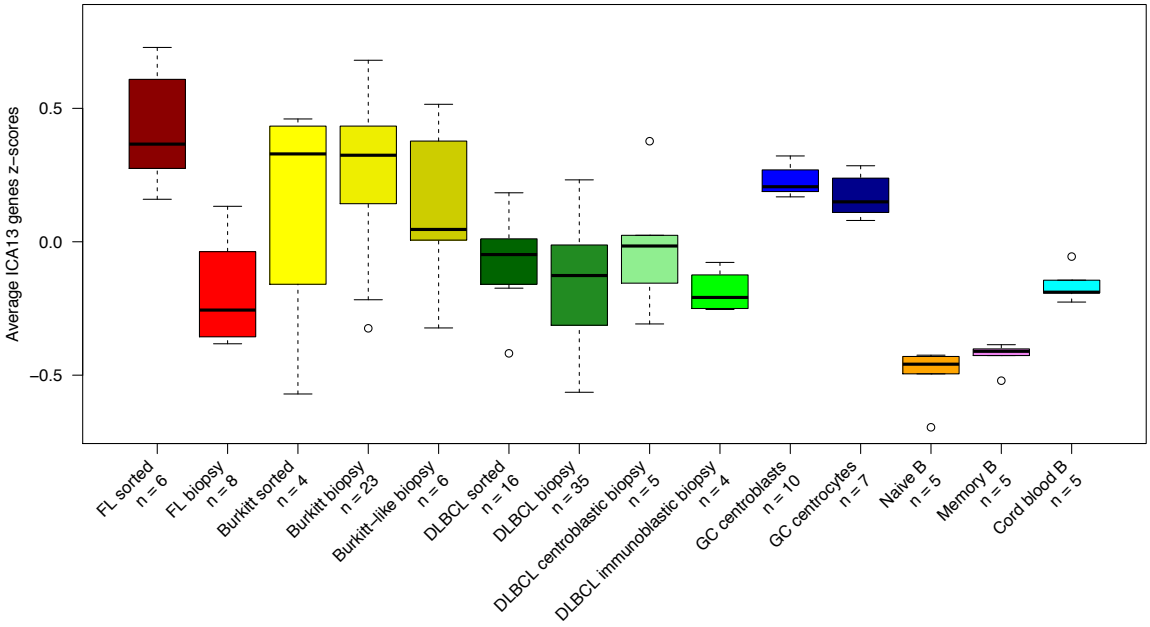


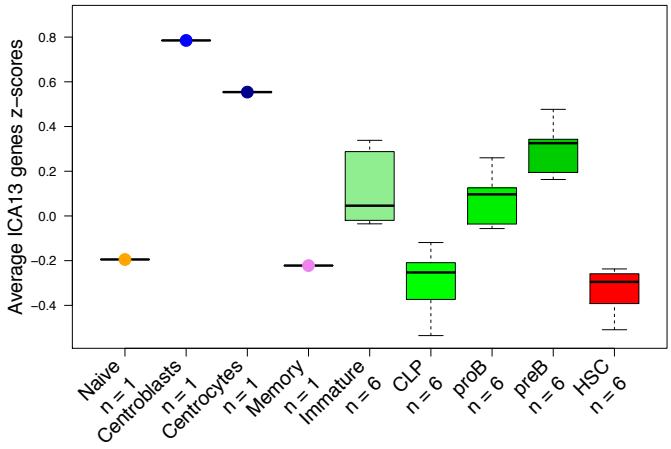
Figure S14: **ICA13 signature in datasets from normal B-cell subsets and primary lymphoma samples.**

The ICA13 signature was assessed in the following GEO datasets: Panel A: GSE2350[13]; B: GSE14714[14]; C: GSE56314[15]. In each dataset, we extracted leading genes from the ICA13 component for which data was available and we computed the average standardized expression of these genes (the negative standardized expression for genes associated with a negative weight in the ICA component). See supplementary methods for details. *Abbreviations: FL: Follicular lymphoma; DLBCL: Diffuse-Large B-Cell Lymphoma; GC: Germinal Center; CLP: Common Lymphoid Progenitor; proB: pro-B-cells; preB: pre-B-cells; HSC: Hematopoietic Stem Cells.*

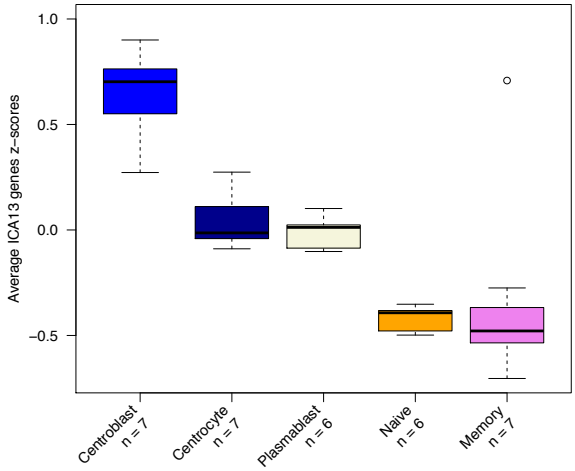
A. ICA13 signature in GSE2350



B. ICA13 signature in GSE14714



C. ICA13 signature in GSE56314



Supplementary Tables

Table S1: Clinical characteristics of patients in the PRIMA training cohort.

In the PRIMA trial, patients diagnosed with high-tumor-burden FL were first treated by one of the three rituximab-chemotherapy regimens (R-CHOP, R-CVP, R-FCM). After this induction phase, patients who obtained a complete or partial response were randomly assigned to observation or rituximab maintenance for 2 years. Clinical characteristics are shown for patients included only in the PFS-supervised analysis (n=134).
Clinical and treatment characteristics were compared using Fisher's exact test. There were no significant differences in the patient clinical characteristics between the whole training cohort (n=160) and all patients in the PRIMA trial (n=1135), except for the frequency of B-symptoms and the induction regimen.
* The whole population from the PRIMA study initially included 1,193 patients. We excluded 58 cases enrolled with FL grade 3B or lymphoma subtypes other than FL. The reference population thus includes 1,135 FL patients.
**Because of rounding, percentages may not total 100.
*** Except grade 3B.
Abbreviations: BM, bone marrow; LDH, lactate dehydrogenase; R, rituximab; CHOP, *cyclophosphamide, doxorubicin, vincristine, prednisone; CVP, cyclophosphamide, vincristine, prednisone; FCM, fludarabine, cyclophosphamide, mitoxantrone; CR, complete response; Cru, unconfirmed CR; PR, partial response; SD, stable disease; PD, progressive disease; NR, not reached.*

Cohort	**Training cohort**	**PRIMA population***	
Number of patients	**N=134**	**N=1135**	
	*no./total no. (%)***	*no./total no. (%)***	*P-value*
Follow-up duration since registration in PRIMA trial			
median, IQR (yrs)	6.6 (6.0-7.0)	6.6 (6.0-7.0)	
Baseline characteristics			
Age > 60 years	47/134 (35)	402/1135 (35)	ns
Male sex	68/134 (51)	590/1135 (52)	ns
Ann Arbor stage III/IV	123/134 (92)	1026/1135 (90)	ns
ECOG PS ≥ 1	43/134 (32)	413/1135 (36)	ns
B symptoms present	30/134 (22)	363/1135 (32)	0.04
BM involvement	83/129 (64)	635/1101 (58)	ns
Elevated LDH	43/134 (32)	378/1130 (33)	ns
Hemoglobin level < 12 g/dL	29/134 (22)	227/1135 (20)	ns
Elevated β2-microglobulin	81/128 (63)	592/1050 (56)	ns
FLIPI score			
0-1 risk factors	23/134 (17)	239/1133 (21)	ns
2 risk factors	53/134 (40)	405/1133 (36)	
3-5 risk factors	58/134 (43)	489/1133 (43)	
Histological grade			
1-2	113/134 (84)	855/994 (86)	
3A	12/134 (9)	77/994 (8)	ns
FL with diffuse area	2/134 (1)	26/994 (3)	
FL of undetermined grade***	7/134 (5)	36/994 (4)	
Induction regimen			
R-CHOP	128/134 (96)	840/1135 (74)	
R-CVP	6/134 (4)	253/1135 (22)	< 0.001
R-FCM	0/134 (0)	42/1135 (4)	
Response to induction regimen			
CR	42/134 (31)	406/1061 (38)	
Cru	49/134 (37)	326/1061 (31)	ns
PR	43/134 (32)	295/1061 (28)	
SD/PD	0/134 (0)	34/1061 (3)	
Maintenance regimen			
Not randomized in PRIMA trial	0/134 (0)	161/1135 (14)	ns
Randomized in PRIMA trial	134/134 (134)	974/1135 (86)	
Observation arm	76/134 (57)	485/974 (50)	ns
Rituximab arm	58/134 (43)	489/974 (50)	
Median progression-free survival since randomisation (yrs)			
Observation arm	4.5	4.1	
Rituximab arm	NR	NR	

Table S2: Genes identified by progression-free survival (PFS)-supervised analysis (n=395).

The "consensus" probe set was selected from the data of the gene-expression array. The column "z-value" indicates the association with PFS: a positive z-value indicates that gene expression is associated with shorter PFS, whereas a negative z-value indicates that gene expression is associated with longer PFS. The columns "NanoString panel" and "Final predictor" indicate which genes were selected for evaluation by NanoString technology (n=95) and were retained in the final 23-gene score, respectively.

Symbol	Description	Probeset	z-value	FDR	Prognostic value confirmed on RNA-seq data	Nanostring panel	Final predictor
METRNL	meteorin, glial cell differentiation regulator-like	225955_at	-5.155	0.0003	yes	yes	yes
ANKRD54	ankyrin repeat domain 54	231847_at	5.456	0.0010	no	no	no
PACS1	phosphofurin acidic cluster sorting protein 1	224658_x_at	-4.442	0.010	no	no	no
VPREB1	pre-B lymphocyte 1	221349_at	5.486	0.010	yes	yes	yes
ACP6	acid phosphatase 6, lysophosphatidic	218795_at	4.569	0.010	yes	yes	no
CNR1	cannabinoid receptor 1 (brain)	1560225_at	4.167	0.010	yes	yes	no
LEPREL4	Synaptonemal complex protein SC65	204078_at	-4.407	0.010	yes	yes	no
TNFRSF1A	tumor necrosis factor receptor superfamily, member 1A	207643_s_at	-4.831	0.010	yes	yes	no
WDR19	WD repeat domain 19	220917_s_at	4.587	0.011	yes	yes	no
SEMA5A	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	213169_at	-4.671	0.012	no	no	no
LOC100506473	uncharacterized LOC100506473	236958_at	4.415	0.012	no	no	no
LCK	lymphocyte-specific protein tyrosine kinase	204890_s_at	-4.387	0.012	yes	yes	no
CAB39L	calcium binding protein 39-like	225915_at	4.784	0.012	yes	yes	no
PIGZ	phosphatidylinositol glycan anchor biosynthesis, class Z	220041_at	4.549	0.012	yes	yes	no
CDHR3	cadherin-related family member 3	235650_at	4.929	0.012	yes	yes	no
PRORSD1P	prolyl-tRNA synthetase associated domain containing 1, pseudogene	237291_at	4.282	0.012	yes	yes	no

PTPN22	protein tyrosine phosphatase, non-receptor type 22 (lymphoid)	208010_s_at	-4.354	0.012	yes	yes	no
TTYH2	tweety homolog 2 (Drosophila)	223741_s_at	-4.560	0.012	yes	yes	no
CDC42EP3	CDC42 effector protein (Rho GTPase binding) 3	209286_at	-4.228	0.012	yes	yes	no
PTPRK	protein tyrosine phosphatase, receptor type, K	203038_at	-4.431	0.012	yes	yes	no
LRRC8C	leucine rich repeat containing 8 family, member C	1558517_s_at	-4.185	0.012	yes	yes	no
ECHDC1	enoyl CoA hydratase domain containing 1	233124_s_at	-4.701	0.012	yes	no	no
MMD	monocyte to macrophage differentiation-associated	244523_at	4.509	0.012	yes	no	no
UNC5B	unc-5 homolog B (C. elegans)	226899_at	-3.977	0.013	no	no	no
CAP2	Adenylyl cyclase-associated protein 2	212554_at	-4.049	0.013	no	no	no
GEMIN7	gem (nuclear organelle) associated protein 7	222821_s_at	4.149	0.013	no	no	no
CXCR4	chemokine (C-X-C motif) receptor 4	209201_x_at	4.021	0.013	yes	yes	yes
ITGB2-AS1	ITGB2 antisense RNA 1	229040_at	-4.213	0.013	yes	yes	no
RNASE6	ribonuclease, RNase A family, k6	213566_at	-4.319	0.013	yes	yes	no
ZDHHC23	zinc finger, DHHC-type containing 23	226912_at	4.143	0.013	yes	yes	no
BEND6	BEN domain containing 6	231175_at	-3.888	0.013	yes	yes	no
MOB3B	MOB kinase activator 3B	226844_at	-4.087	0.013	yes	yes	no
COMMD8	COMM domain containing 8	218351_at	-4.612	0.013	yes	yes	no
LOC100127983	uncharacterized LOC100127983	228107_at	4.308	0.013	yes	no	no
FBXO18	F-box protein, helicase, 18	224683_at	4.280	0.013	yes	no	no
ANKRD36BP2	ankyrin repeat domain 36B pseudogene 2	1556183_at	4.183	0.013	yes	no	no
SNAI3	snail family zinc finger 3	1560228_at	-4.061	0.013	yes	no	no
ANTXR1	anthrax toxin receptor 1	224694_at	-4.181	0.013	yes	no	no
PARD3B	par-3 partitioning defective 3 homolog B (C. elegans)	228411_at	-4.063	0.013	yes	no	no
REPIN1	replication initiator 1	219041_s_at	-4.019	0.015	yes	no	no
STAT4	signal transducer and activator of transcription 4	206118_at	-4.121	0.015	yes	yes	no

INPP5A	inositol polyphosphate-5-phosphatase, 40kDa	203006_at	-3.955	0.015	yes	yes	no
AFF3	AF4/FMR2 family, member 3	227198_at	3.762	0.015	yes	yes	yes
PIGM	phosphatidylinositol glycan anchor biosynthesis, class M	235168_at	4.038	0.015	yes	yes	no
FLT3	fms-related tyrosine kinase 3	206674_at	-3.827	0.016	yes	yes	no
ABRACL	ABRA C-terminal like	223361_at	-4.206	0.016	yes	no	no
DSC2	desmocollin 2	226817_at	-4.164	0.017	yes	yes	no
FAM105A	family with sequence similarity 105, member A	219694_at	-4.217	0.017	yes	no	no
SAMD4A	sterile alpha motif domain containing 4A	212845_at	-4.055	0.017	yes	no	no
RNF144A	Probable E3 ubiquitin-protein ligase RNF144A	204040_at	-3.912	0.018	yes	yes	no
CADPS2	Ca++-dependent secretion activator 2	219572_at	-3.985	0.018	yes	no	no
TAGAP	T-cell activation RhoGTPase activating protein	1552542_s_at	4.050	0.018	yes	yes	yes
HELZ2	helicase with zinc finger 2, transcriptional coactivator	228230_at	-3.704	0.018	no	no	no
SHISA8	shisa homolog 8 (Xenopus laevis)	244467_at	-3.804	0.018	yes	yes	yes
LGALS2	lectin, galactoside-binding, soluble, 2	208450_at	-4.325	0.018	yes	yes	no
IGSF10	immunoglobulin superfamily, member 10	1556579_s_at	-3.851	0.018	yes	yes	no
RCBTB1	regulator of chromosome condensation (RCC1) and BTB (POZ) domain containing protein 1	218352_at	3.949	0.018	yes	no	no
SHB	Src homology 2 domain containing adaptor protein B	1557458_s_at	-3.978	0.018	yes	no	no
CCAR2=KIAA1967	KIAA1967/DBC1/Cell cycle and apoptosis regulator protein 2	225193_at	4.046	0.019	no	no	no
KISS1R	KISS1 receptor	242517_at	-3.550	0.019	no	no	no
WNK2	WNK lysine deficient protein kinase 2	227217_at	4.347	0.019	yes	no	no
RASA3	RAS p21 protein activator 3	225562_at	-3.762	0.019	yes	no	no
IP6K2	inositol hexakisphosphate kinase 2	223165_s_at	3.734	0.019	no	no	no
ZNF267	zinc finger protein 267	219540_at	-3.927	0.019	no	no	no

PMS1	PMS1 postmeiotic segregation increased 1 (S. cerevisiae)/ DNA mismatch repair protein PMS1	236609_at	3.850	0.019	no	no	no
DCAF12	DDB1 and CUL4 associated factor 12	224789_at	3.984	0.019	yes	yes	yes
CSNK1E	casein kinase 1, epsilon	202332_at	4.083	0.019	yes	yes	no
AMICA1	Junctional adhesion molecule-like / adhesion molecule, interacts with CXADR antigen 1	228094_at	-4.068	0.019	yes	yes	no
IL15	interleukin 15	217371_s_at	-4.081	0.019	yes	yes	no
ZNF677	zinc finger protein 677	228974_at	-3.837	0.019	yes	yes	no
TIGD1	tigger transposable element derived 1	1553099_at	3.793	0.019	yes	no	no
MOAP1	modulator of apoptosis 1	212508_at	3.701	0.019	yes	no	no
PGBD1	piggyBac transposable element derived 1	235411_at	3.995	0.019	yes	no	no
CLEC4F	C-type lectin domain family 4, member F	1552410_at	-3.622	0.019	yes	no	no
MEIS3P1	Putative homeobox protein Meis3-like 1	214077_x_at	-4.009	0.019	no	no	no
DUSP7	dual specificity phosphatase 7	213848_at	-3.935	0.019	yes	no	no
GAS2L1	growth arrest-specific 2 like 1	31874_at	-3.871	0.020	yes	no	no
NAV1	neuron navigator 1	224772_at	-3.890	0.020	yes	yes	no
DDO	D-aspartate oxidase	222134_at	-3.740	0.020	yes	no	no
CHST10	carbohydrate sulfotransferase 10	204065_at	3.879	0.020	yes	no	no
ABCC5	ATP-binding cassette, sub-family C (CFTR/MRP), member 5	226363_at	4.133	0.020	yes	no	no
OCRL	oculocerebrorenal syndrome of Lowe/ Inositol polyphosphate 5-phosphatase	203446_s_at	3.854	0.020	yes	no	no
H2AFY2	H2A histone family, member Y2	218445_at	-3.771	0.020	yes	no	no
MBTPS1	membrane-bound transcription factor peptidase, site 1	217543_s_at	3.833	0.020	yes	no	no
MIPOL1	mirror-image polydactyly 1	244246_at	-3.623	0.021	no	no	no
TTC21A	tetratricopeptide repeat domain 21A	230997_at	4.019	0.021	yes	no	no
ZFP112	zinc finger protein 112 homolog (mouse)	222237_s_at	-3.739	0.021	no	no	no

VCL	vinculin	200931_s_at	-3.642	0.021	yes	yes	yes
TLR2	toll-like receptor 2	204924_at	-3.926	0.021	yes	yes	no
USP13	ubiquitin specific peptidase 13 (isopeptidase T-3)	205356_at	3.658	0.021	yes	yes	no
F2R	coagulation factor II (thrombin) receptor	203989_x_at	-3.724	0.021	yes	yes	no
RAD51-AS1	RAD51 antisense RNA 1 (head to head)	242518_at	3.890	0.021	yes	no	no
PPDPF	pancreatic progenitor cell differentiation and proliferation factor homolog (zebrafish)	233571_x_at	-3.798	0.021	yes	no	no
NACC2	NACC family member 2, BEN and BTB (POZ) domain containing	212993_at	-3.957	0.021	yes	no	no
RXRA	retinoid X receptor, alpha	202449_s_at	-3.850	0.021	yes	no	no
CCDC89	coiled-coil domain containing 89	1553228_at	-3.621	0.021	no	no	no
DPYD	dihydropyrimidine dehydrogenase	204646_at	-3.884	0.021	no	no	no
TYROBP=DAP12	TYRO protein tyrosine kinase binding protein	204122_at	-3.931	0.021	yes	yes	no
FAM46A	family with sequence similarity 46, member A	224973_at	-3.921	0.021	yes	yes	no
CCND1	cyclin D1	208712_at	-3.788	0.021	yes	yes	no
RHOU	ras homolog family member U	223168_at	-3.887	0.021	yes	yes	no
MGAT4A	mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isozyme A	219797_at	-3.822	0.021	yes	yes	no
STK33	serine/threonine kinase 33	228086_at	4.291	0.021	yes	no	no
B3GNTL1	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase-like 1	213589_s_at	-3.685	0.021	yes	no	no
CYB561	cytochrome b561	209163_at	-3.820	0.021	yes	no	no
ZNF578	zinc finger protein 578	1562245_a_at	-3.521	0.022	no	no	no
ASAP2	ArfGAP with SH3 domain, ankyrin repeat and PH domain 2	206414_s_at	-3.832	0.023	no	no	no
EMP2	epithelial membrane protein 2	225078_at	-3.888	0.023	yes	yes	no
EPB41L3	erythrocyte membrane protein band 4.1-like 3	206710_s_at	-3.979	0.023	yes	yes	no
TNFSF13 = APRIL	tumor necrosis factor (ligand)	210314_x_at	-4.082	0.023	yes	yes	no

	superfamily, member 13						
ARPC2	actin related protein 2/3 complex, subunit 2, 34kDa	207988_s_at	-3.846	0.023	yes	no	no
GADD45A	growth arrest and DNA-damage-inducible, alpha	203725_at	4.013	0.023	yes	yes	yes
ARHGEF40	Rho guanine nucleotide exchange factor (GEF) 40	58780_s_at	-3.684	0.023	yes	no	no
KCNS3	potassium voltage-gated channel, delayed-rectifier, subfamily S, member 3	205968_at	-3.467	0.023	yes	no	no
C6orf99	chromosome 6 open reading frame 99	230487_at	3.956	0.024	yes	no	no
HUWE1	HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase	236294_at	3.843	0.024	yes	no	no
RAB11FIP5	RAB11 family interacting protein 5 (class I)	210879_s_at	-3.737	0.024	yes	no	no
MLK4	KIAA1804/mixed lineage kinase 4	228565_at	4.145	0.025	no	no	no
PDE4B	phosphodiesterase 4B, cAMP-specific	211302_s_at	3.623	0.025	yes	yes	no
TAPT1	transmembrane anterior posterior transformation 1	238798_at	4.104	0.025	yes	no	no
ASPH	aspartate beta-hydroxylase	210896_s_at	-3.755	0.026	no	no	no
PIK3R3	phosphoinositide-3-kinase, regulatory subunit 3 (gamma)	202743_at	-3.612	0.026	no	no	no
PPP2R3A	protein phosphatase 2A, regulatory subunit B", alpha	209632_at	-3.612	0.026	no	no	no
ZNF585A	zinc finger protein 585A	1554824_at	3.699	0.026	yes	no	no
C15orf48 = NMES1	chromosome 15 open reading frame 48	223484_at	-3.475	0.027	yes	yes	no
LOC100270804	uncharacterized LOC100270804	229722_at	3.787	0.027	no	no	no
MGC21881	uncharacterized locus MGC21881/ Putative UPF0633 protein MGC21881	228040_at	3.520	0.027	yes	no	no
ZNRF2	E3 ubiquitin-protein ligase ZNRF2	226261_at	-3.617	0.027	yes	no	no
TP53I13	tumor protein p53 inducible protein 13	225605_at	3.774	0.027	no	no	no
FAM171A1	family with sequence similarity 171, member A1	212771_at	-3.668	0.027	yes	no	no
PRKAG2-AS1	PRKAG2 antisense RNA 1	229156_s_at	-3.619	0.027	yes	no	no

SPIN4	spindlin family, member 4	228654_at	3.816	0.027	yes	no	no
ARHGEF10L	Rho guanine nucleotide exchange factor (GEF) 10-like	221656_s_at	-3.589	0.027	yes	no	no
CSTA	cystatin A (stefin A)	204971_at	-3.860	0.028	yes	yes	no
PDE10A	phosphodiesterase 10A	205501_at	-3.749	0.028	no	no	no
ZHX1/C8ORF76	ZHX1-C8ORF76 readthrough	1554063_at	3.624	0.028	no	no	no
ZNF8	zinc finger protein 8	214901_at	3.637	0.028	no	no	no
SRSF8	serine/arginine-rich splicing factor 8	238929_at	3.615	0.028	no	no	no
PEX5	peroxisomal biogenesis factor 5	203244_at	3.867	0.028	yes	yes	no
GNAQ	guanine nucleotide binding protein (G protein), q polypeptide	224862_at	-3.777	0.028	yes	yes	no
SOX2	SRY (sex determining region Y)-box 2	228038_at	4.010	0.028	yes	yes	no
RIN2	Ras and Rab interactor 2	209684_at	-3.746	0.028	yes	no	no
UHRF2	ubiquitin-like with PHD and ring finger domains 2, E3 ubiquitin protein ligase	225610_at	3.669	0.028	yes	no	no
GNAS-AS1	GNAS antisense RNA 1	232881_at	3.881	0.028	yes	no	no
COG1	component of oligomeric golgi complex 1	227784_s_at	3.697	0.028	yes	no	no
SH3RF1=POSH	E3 ubiquitin-protein ligase SH3RF1	225589_at	-3.488	0.028	yes	no	no
TMEM255A	transmembrane protein 255A	219895_at	-3.765	0.028	yes	no	no
KIAA0040	KIAA0040	203143_s_at	3.611	0.028	yes	yes	yes
MCC	mutated in colorectal cancers	226225_at	-3.808	0.028	no	no	no
ALDH2	aldehyde dehydrogenase 2 family (mitochondrial)	201425_at	-3.564	0.028	yes	yes	yes
LOC285972	uncharacterized LOC285972	1560762_at	-3.590	0.028	yes	no	no
PNN	pinin, desmosome associated protein	1567213_at	3.581	0.028	yes	no	no
RGMA	RGM domain family, member A	223468_s_at	-3.487	0.028	yes	no	no
TTC39B	tetratricopeptide repeat domain 39B	242477_at	-3.597	0.029	yes	yes	no
UBAP2	ubiquitin associated protein 2	221839_s_at	3.517	0.030	yes	no	no
PLCB4	phospholipase C, beta 4	203895_at	-3.600	0.030	no	no	no
LOC100287525	uncharacterized LOC100287525	230606_at	3.750	0.030	no	no	no

ITM2C	integral membrane protein 2C	221004_s_at	-3.734	0.030	yes	yes	no
GUCD1	guanylyl cyclase domain containing 1	223039_at	3.533	0.030	yes	no	no
SHE	Src homology 2 domain containing E	229910_at	-3.594	0.030	no	no	no
TLR5	toll-like receptor 5	210166_at	-3.435	0.030	yes	yes	no
LSAMP	limbic system-associated membrane protein	228218_at	-3.556	0.031	yes	no	no
LOC100130458	uncharacterized LOC100130458	239214_at	3.289	0.031	no	no	no
WDFY3 = ALFY	WD repeat and FYVE domain containing 3	212606_at	-3.698	0.031	no	no	no
ZAK	sterile alpha motif and leucine zipper containing kinase AZK/Mitogen-activated protein kinase kinase kinase MLT	222757_s_at	3.497	0.031	no	no	no
USP44	ubiquitin specific peptidase 44	224048_at	3.568	0.031	yes	yes	yes
TYMP	thymidine phosphorylase	217497_at	-3.597	0.031	yes	yes	no
PRDM15 = ZNF298	PR domain containing 15	231931_at	3.404	0.031	yes	yes	yes
CCR2	chemokine (C-C motif) receptor 2	206978_at	-3.627	0.031	yes	yes	no
CD79B	CD79b molecule, immunoglobulin-associated beta	205297_s_at	3.374	0.031	yes	yes	no
SH2D3A = NSP1	SH2 domain containing 3A	219513_s_at	-3.448	0.031	yes	no	no
EFHC1	EF-hand domain (C-terminal) containing 1	225656_at	3.624	0.032	yes	yes	no
LOC645513	uncharacterized LOC645513	1561759_at	3.693	0.032	no	no	no
EPS8L2	EPS8-like 2	218180_s_at	-3.325	0.032	yes	no	no
CLIC2	chloride intracellular channel 2	213415_at	-3.748	0.032	yes	no	no
ZSWIM8-AS1	ZSWIM8 antisense RNA 1	214867_at	3.622	0.033	no	no	no
NFX1	nuclear transcription factor, X-box binding 1	1553348_a_at	3.475	0.033	yes	yes	no
SEMA4B	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4B	234725_s_at	3.597	0.033	yes	yes	yes
KIAA0753	KIAA0753	204711_at	3.825	0.033	yes	no	no
PWWP2B	PWWP domain containing 2B	227999_at	-3.193	0.034	yes	no	no
NUP214	nucleoporin 214kDa	228902_at	3.603	0.034	yes	no	no

RGS10	regulator of G-protein signaling 10	204319_s_at	-3.454	0.034	yes	yes	yes
FCRL2	Fc receptor-like 2	221239_s_at	3.246	0.034	yes	yes	yes
FRMD4B	FERM domain containing 4B	213056_at	-3.666	0.035	no	no	no
MDM4	Mdm4 p53 binding protein homolog (mouse)	225742_at	3.445	0.035	yes	yes	no
C11orf74	chromosome 11 open reading frame 74	228249_at	-3.645	0.035	yes	no	no
NRARP	NOTCH-regulated ankyrin repeat protein	226499_at	-3.420	0.035	yes	no	no
ADAMTS3	ADAM metallopeptidase with thrombospondin type 1 motif, 3	214913_at	-3.552	0.035	yes	no	no
DYNLT1	dynein, light chain, Tctex-type 1	201999_s_at	-3.637	0.035	yes	no	no
SPTLC3	serine palmitoyltransferase, long chain base subunit 3	227752_at	-3.177	0.035	no	no	no
SLC2A10	solute carrier family 2 (facilitated glucose transporter), member 10	221024_s_at	-3.418	0.035	no	no	no
SAP30BP	SAP30 binding protein	222475_at	3.493	0.035	no	no	no
EXPH5	exophilin 5	213929_at	-3.236	0.035	no	no	no
TJP1	tight junction protein 1	202011_at	-3.521	0.035	no	no	no
SSPN	sarcospan	226932_at	-3.504	0.035	no	no	no
LOC100509635	uncharacterized LOC100509635	226865_at	-3.399	0.035	no	no	no
EML6	echinoderm microtubule associated protein like 6	229656_s_at	3.024	0.035	yes	yes	yes
RASSF6	Ras association (RalGDS/AF-6) domain family member 6	229147_at	3.317	0.035	yes	yes	yes
ICA1L	islet cell autoantigen 1,69kDa-like	230454_at	3.576	0.035	yes	no	no
ODF2	outer dense fiber of sperm tails 2	225617_at	3.619	0.035	yes	no	no
PLCXD1	phosphatidylinositol-specific phospholipase C, X domain containing 1	218951_s_at	3.536	0.035	yes	no	no
HIST1H2BC	histone cluster 1, H2bc	236193_at	3.713	0.035	yes	no	no
BNC2	basonuclin 2	230722_at	-3.336	0.035	no	no	no
LRP10	low density lipoprotein receptor-related protein 10	231861_at	3.837	0.035	no	no	no
FGF2	fibroblast growth factor 2 (basic)	204422_s_at	-3.517	0.036	no	no	no

VN1R1	vomeronasal 1 receptor 1	221412_at	3.549	0.036	no	no	no
PQLC3	PQ loop repeat containing 3	225579_at	-3.472	0.036	yes	no	no
TMEM187	transmembrane protein 187	204340_at	-3.515	0.036	yes	no	no
GALNT11	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 11 (GalNAc-T11)	219013_at	-3.527	0.036	yes	no	no
L3MBTL1	Lethal(3)malignant brain tumor-like protein 1	210306_at	3.384	0.036	yes	no	no
LILRA4	leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 4	210313_at	-3.190	0.036	yes	no	no
ALDH5A1	aldehyde dehydrogenase 5 family, member A1	203609_s_at	3.655	0.037	yes	yes	no
IFT172	intraflagellar transport 172 homolog (Chlamydomonas)	226324_s_at	3.575	0.037	yes	no	no
KCNQ1OT1	KCNQ1 opposite strand/antisense transcript 1 (non-protein coding)	243435_at	3.996	0.037	yes	no	no
PLEKHM1	pleckstrin homology domain containing, family M (with RUN domain) member 1	212717_at	3.528	0.037	yes	no	no
PRELP	proline/arginine-rich end leucine-rich repeat protein	228224_at	-3.129	0.037	yes	no	no
UBASH3B	ubiquitin associated and SH3 domain containing B	238462_at	-3.540	0.037	yes	no	no
SPG20	spastic paraplegia 20 (Troyer syndrome)	236600_at	-3.473	0.037	yes	no	no
EHD4	EH-domain containing 4	229074_at	-3.415	0.037	yes	no	no
NTNG1	netrin G1	236088_at	-3.033	0.037	yes	no	no
GH1	growth hormone 1/Somatotropin	206886_x_at	3.555	0.037	yes	no	no
CCDC136/NAG6	coiled-coil domain containing 136	226972_s_at	3.655	0.037	yes	no	no
ZNF451	zinc finger protein 451	215012_at	3.549	0.037	yes	no	no
VASP	vasodilator-stimulated phosphoprotein	202205_at	-3.565	0.038	yes	no	no
FUBP1	far upstream element (FUSE) binding protein 1	214094_at	3.436	0.038	no	no	no
MYCL1	v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived (avian)	214058_at	-3.372	0.038	yes	no	no
TSPAN6	tetraspanin 6	209109_s_at	-3.484	0.039	yes	yes	no

C1orf220	chromosome 1 open reading frame 220	1563876_at	3.561	0.039	yes	no	no
DHTKD1	dehydrogenase E1 and transketolase domain containing 1	227094_at	3.421	0.039	yes	no	no
PLEKHG4	pleckstrin homology domain containing, family G (with RhoGef domain) member 4	228171_s_at	-3.344	0.039	no	no	no
ZSCAN29	zinc finger and SCAN domain containing 29	226562_at	3.305	0.039	yes	no	no
MAGI2-AS3	MAGI2 antisense RNA 3	227554_at	-3.372	0.039	no	no	no
LOC100996286	uncharacterized LOC100996286	1559205_s_at	-3.317	0.039	no	no	no
ATRIP	ATR interacting protein	1552937_s_at	3.494	0.039	no	no	no
RNASEH2C	ribonuclease H2, subunit C	227543_at	3.361	0.039	no	no	no
TCF4	transcription factor 4	228837_at	3.446	0.039	yes	yes	yes
GLRX	glutaredoxin (thioltransferase)	209276_s_at	-3.578	0.039	yes	no	no
PRKCE	protein kinase C, epsilon	226101_at	3.166	0.039	yes	no	no
ZDHHC9	zinc finger, DHHC-type containing 9	222451_s_at	3.558	0.039	yes	no	no
HIST1H2AC	histone cluster 1, H2ac	215071_s_at	3.359	0.039	yes	no	no
SLC8A1	solute carrier family 8 (sodium/calcium exchanger), member 1	235518_at	-3.505	0.039	no	no	no
TMEM45A	transmembrane protein 45A	219410_at	-3.254	0.039	no	no	no
LOC100128288	uncharacterized LOC100128288	1559045_at	3.343	0.039	no	no	no
CYTH3	cytohesin 3	225147_at	3.638	0.039	yes	no	no
SYTL4	synaptotagmin-like 4	227703_s_at	-3.014	0.039	yes	no	no
FOXP1-IT1	FOXP1 intronic transcript 1 (non-protein coding)	232096_x_at	3.625	0.039	yes	no	no
KIAA0895	KIAA0895	213424_at	3.958	0.039	yes	no	no
KIAA1598	KIAA1598	221802_s_at	-3.514	0.039	yes	no	no
PTPRS	protein tyrosine phosphatase, receptor type, S	226571_s_at	-3.278	0.039	yes	no	no
SPRY2	sprouty homolog 2 (Drosophila)	204011_at	-3.617	0.040	yes	no	no
ITGA1	integrin, alpha 1	214660_at	-3.317	0.040	yes	no	no
BMP3	bone morphogenetic protein 3	208244_at	3.890	0.040	yes	no	no
PLXNA4	plexin A4	232317_at	-3.199	0.040	yes	no	no

HADH	hydroxyacyl-CoA dehydrogenase	201035_s_at	3.359	0.040	yes	no	no
TBC1D14	TBC1 domain family, member 14	224622_at	-3.337	0.040	no	no	no
LOC100129502	uncharacterized LOC100129502	228791_at		0.040	no	no	no
RAB7B	RAB7B, member RAS oncogene family	230266_at	-3.301	0.040	no	no	no
FOXO1	forkhead box O1	202723_s_at	3.460	0.040	yes	yes	yes
JAM3	junctional adhesion molecule 3	212813_at	-3.383	0.040	yes	yes	no
CLEC4C	C-type lectin domain family 4, member C	1552552_s_at	-3.069	0.040	yes	no	no
FAM214A	family with sequence similarity 214, member A	225327_at	3.285	0.040	yes	no	no
CHSY1	chondroitin sulfate synthase 1	203044_at	-3.436	0.041	no	no	no
PPP4R1	protein phosphatase 4, regulatory subunit 1	201594_s_at	3.402	0.041	yes	no	no
CXCR3	chemokine (C-X-C motif) receptor 3	207681_at	-3.323	0.041	yes	no	no
ANXA1	annexin A1	201012_at	-3.450	0.041	yes	no	no
CSF1R	colony stimulating factor 1 receptor	203104_at	-3.351	0.041	yes	no	no
SLC1A3	solute carrier family 1 (glial high affinity glutamate transporter), member 3	202800_at	-3.284	0.041	yes	no	no
ARL5B-AS1	ARL5B antisense RNA 1	242239_at	3.374	0.041	no	no	no
TRIB2	tribbles homolog 2 (Drosophila)	202479_s_at	3.238	0.041	yes	yes	no
HCST=DAP10	hematopoietic cell signal transducer	223640_at	-3.434	0.041	yes	yes	no
CRY1	cryptochrome 1 (photolyase-like)	209674_at	-3.498	0.041	yes	yes	no
GNAZ	guanine nucleotide binding protein (G protein), alpha z polypeptide	204993_at	3.413	0.041	yes	no	no
BAALC	brain and acute leukemia, cytoplasmic	218899_s_at	-3.201	0.041	yes	no	no
ADARB1	adenosine deaminase, RNA-specific, B1	203865_s_at	-3.207	0.041	yes	no	no
PRICKLE1	prickle homolog 1 (Drosophila)	230708_at	-3.264	0.041	yes	no	no
MAP9	microtubule-associated protein 9	220145_at	3.298	0.041	yes	no	no

PSMD11	proteasome (prosome, macropain) 26S subunit, non-ATPase, 11	208776_at	3.433	0.041	yes	no	no
CGNL1	cingulin-like 1	225817_at	-3.164	0.041	yes	no	no
PDDC1	Parkinson disease 7 domain containing 1	227968_at	3.317	0.041	yes	no	no
ST6GALNAC1	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1	227725_at	-3.223	0.041	yes	no	no
WDR13	WD repeat domain 13	222138_s_at	3.544	0.041	yes	no	no
TMEM133	transmembrane protein 133	223595_at	-3.446	0.041	no	no	no
TAF10	TAF10 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 30kDa	235906_at	3.565	0.041	no	no	no
E2F5	E2F transcription factor 5, p130-binding	221586_s_at	3.211	0.041	yes	yes	yes
MAN1C1	mannosidase, alpha, class 1C, member 1	218918_at	-3.441	0.041	yes	no	no
CST7	cystatin F (leukocystatin)	210140_at	-3.469	0.041	yes	no	no
RAB32	RAB32, member RAS oncogene family	204214_s_at	-3.649	0.041	yes	no	no
EGLN3	egl nine homolog 3 (C. elegans)	219232_s_at	-3.386	0.041	yes	no	no
TIMP2	TIMP metallopeptidase inhibitor 2	224560_at	-3.335	0.041	yes	no	no
RPS12	ribosomal protein S12	213377_x_at	-3.648	0.041	yes	no	no
SLCO2B1	solute carrier organic anion transporter family, member 2B1	203473_at	-3.519	0.041	yes	no	no
TCP11L2	t-complex 11, testis-specific-like 2	1559413_at	3.287	0.041	yes	no	no
FZD5	frizzled family receptor 5	221245_s_at	-3.310	0.041	yes	no	no
DCHS1	Protocadherin-16	222101_s_at	-3.338	0.041	no	no	no
IRF5	interferon regulatory factor 5	239412_at	3.304	0.041	no	no	no
DPY19L2P2	dpy-19-like 2 pseudogene 2 (C. elegans)/ Putative C-mannosyltransferase DPY19L2P2	215143_at	3.332	0.042	yes	no	no
MAN2A1	mannosidase, alpha, class 2A, member 1	235103_at	-3.348	0.042	yes	no	no
LINC00883	long intergenic non-protein coding RNA 883	235606_at	3.373	0.042	yes	no	no

EPHA1-AS1	EPHA1 antisense RNA 1	1562249_at	-3.085	0.042	no	no	no
SOCS7	suppressor of cytokine signaling 7	228662_at	3.263	0.042	no	no	no
ANKS6	ankyrin repeat and sterile alpha motif domain containing 6	226586_at	3.492	0.042	yes	no	no
LHFPL2	lipoma HMGIC fusion partner-like 2	212658_at	3.078	0.042	yes	no	no
SLC24A3	solute carrier family 24 (sodium/potassium/calcium exchanger), member 3	219090_at	-3.166	0.042	yes	no	no
CAP1	CAP, adenylate cyclase-associated protein 1 (yeast)	200625_s_at	-3.273	0.042	yes	no	no
BICC1	bicaudal C homolog 1 (Drosophila)	213429_at	-3.250	0.042	yes	no	no
DICER1-AS1	DICER1 antisense RNA 1	229227_at	3.620	0.042	yes	no	no
DNM3OS	DNM3 opposite strand/antisense RNA	232090_at	-3.172	0.042	yes	no	no
ORAI2	ORAI calcium release-activated calcium modulator 2	217529_at	3.258	0.043	yes	yes	yes
RAG2	recombination activating gene 2	215117_at	3.566	0.043	yes	no	no
LOC100506469	uncharacterized LOC100506469	222280_at	3.219	0.043	yes	no	no
FAM120AOS	family with sequence similarity 120A opposite strand	239391_at	3.335	0.043	no	no	no
ACSL6	acyl-CoA synthetase long-chain family member 6	229725_at	-3.399	0.043	no	no	no
RLN1	relaxin 1	211753_s_at	3.459	0.043	no	no	no
MS4A6A	membrane-spanning 4-domains, subfamily A, member 6A	223280_x_at	-3.503	0.043	yes	yes	no
LST1	leukocyte specific transcript 1	214181_x_at	-3.388	0.043	yes	yes	no
RNF130	ring finger protein 130	217865_at	-3.481	0.043	yes	no	no
NPC2	Niemann-Pick disease, type C2	200701_at	-3.345	0.043	yes	no	no
RTN1	reticulon 1	203485_at	-3.432	0.043	yes	no	no
RGP1	RGP1 retrograde golgi transport homolog (S. cerevisiae)	203169_at	3.323	0.043	yes	no	no
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	209993_at	-3.147	0.043	yes	yes	yes
FCER1G	Fc fragment of IgE, high affinity I, receptor for; gamma	1554899_s_at	-3.369	0.043	yes	yes	no

	polypeptide						
XK	X-linked Kx blood group (McLeod syndrome)	206698_at	3.714	0.043	yes	no	no
ETV7	ets variant 7	224225_s_at	-2.999	0.043	yes	no	no
LOC100506965	uncharacterized LOC100506965	226587_at	3.214	0.043	yes	no	no
ITGAX	integrin, alpha X (complement component 3 receptor 4 subunit)	210184_at	-3.224	0.043	yes	no	no
GZMM	granzyme M (lymphocyte met-ase 1)	207460_at	-3.251	0.043	yes	no	no
ASB16-AS1	ASB16 antisense RNA 1	1555981_at	3.416	0.043	yes	no	no
FAT4	FAT atypical cadherin 4	219427_at	-3.337	0.043	no	no	no
SNRNP35	small nuclear ribonucleoprotein 35kDa (U11/U12)	205300_s_at	3.272	0.043	no	no	no
N4BP2L2-IT2	N4BPL2 intronic transcript 2 (non-protein coding)	215105_at	3.425	0.043	no	no	no
TPBG	trophoblast glycoprotein	203476_at	-3.130	0.043	yes	no	no
DHRS3	dehydrogenase/reductase (SDR family) member 3	202481_at	-3.297	0.043	yes	no	no
METRN	meteorin, glial cell differentiation regulator	232269_x_at	-3.233	0.043	yes	no	no
MEX3C	RNA-binding E3 ubiquitin-protein ligase MEX3C	222567_s_at	3.499	0.043	yes	no	no
SH2B3	SH2B adaptor protein 3	203320_at	-3.314	0.043	no	no	no
EPSTI1	epithelial stromal interaction 1 (breast)	227609_at	-3.459	0.043	yes	no	no
FAM13A	family with sequence similarity 13, member A	217047_s_at	-3.385	0.044	no	no	no
LOC100507018	uncharacterized LOC100507018	227745_at	3.265	0.044	no	no	no
RGS6	regulator of G-protein signaling 6	210270_at	3.473	0.044	yes	no	no
ARHGAP28	Rho GTPase activating protein 28	227911_at	-3.066	0.044	no	no	no
GLIPR2	GLI pathogenesis-related 2	225604_s_at	-3.270	0.044	yes	yes	no
OGT	O-linked N-acetylglucosamine (GlcNAc) transferase	207563_s_at	3.295	0.044	yes	yes	no
GSTK1	glutathione S-transferase kappa 1	217751_at	-3.250	0.044	yes	no	no
BCL2A1	BCL2-related protein A1	205681_at	-3.286	0.044	yes	no	no

C19orf18	chromosome 19 open reading frame 18	236847_at	3.404	0.044	yes	no	no
NCF2	neutrophil cytosolic factor 2	209949_at	-3.211	0.044	yes	no	no
ST3GAL6	ST3 beta-galactoside alpha-2,3-sialyltransferase 6	213355_at	-3.138	0.044	yes	no	no
TMEM204	transmembrane protein 204	219315_s_at	-3.332	0.044	yes	no	no
RUNDC1	RUN domain containing 1	235040_at	3.314	0.044	yes	no	no
DTX4	deltex homolog 4 (Drosophila)	212611_at	3.461	0.044	yes	no	no
PHF2	PHD finger protein 2	212726_at	3.256	0.044	yes	no	no
PAX5	paired box 5	221969_at	3.064	0.044	yes	no	no
PPAT	phosphoribosyl pyrophosphate amidotransferase	209433_s_at	2.999	0.044	no	no	no
GABRB2	gamma-aminobutyric acid (GABA) A receptor, beta 2	242344_at	3.349	0.044	yes	no	no
LOC644656	uncharacterized LOC644656	229870_at	3.148	0.045	no	no	no
EGFR	epidermal growth factor receptor	224999_at	-3.255	0.045	no	no	no
TANC2	tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 2	224952_at	-3.056	0.045	no	no	no
FLJ20021	uncharacterized LOC90024	228832_at	3.259	0.046	no	no	no
TET1	tet methylcytosine dioxygenase 1	228906_at	3.295	0.046	yes	no	no
PRKCDBP	protein kinase C, delta binding protein	213010_at	-3.269	0.046	yes	no	no
PLXDC1	plexin domain containing 1	219700_at	-3.176	0.046	yes	no	no
C10orf76	chromosome 10 open reading frame 76	55662_at	-3.257	0.046	no	no	no
IL12RB1	interleukin 12 receptor, beta 1	1552584_at	-3.205	0.046	yes	no	no
CCDC117	coiled-coil domain containing 117	225644_at	3.208	0.046	yes	no	no
DUSP8	dual specificity phosphatase 8	206374_at	-3.124	0.047	yes	no	no
SPTA1	spectrin, alpha, erythrocytic 1 (elliptocytosis 2)	206937_at	3.740	0.047	yes	no	no
INF2	inverted formin, FH2 and WH2 domain containing	224469_s_at	-3.147	0.047	yes	no	no
C3orf52/TTMP	TPA-induced transmembrane protein/chromosome 3 open reading frame 52	219474_at	3.663	0.047	yes	no	no
MYO1F	myosin IF	213733_at	-3.151	0.047	yes	no	no

NT5E	5'-nucleotidase, ecto (CD73)	203939_at	-3.371	0.047	yes	no	no
ZMIZ1	zinc finger, MIZ-type containing 1	212124_at	-3.239	0.047	yes	no	no
PIWIL4	piwi-like RNA-mediated gene silencing 4	230480_at	-3.226	0.047	yes	no	no
RASGRF2	Ras protein-specific guanine nucleotide-releasing factor 2	228109_at	-3.135	0.047	no	no	no
CACNA1A	calcium channel, voltage-dependent, P/Q type, alpha 1A subunit	214933_at	-3.087	0.047	yes	no	no
GZMB	granzyme B (granzyme 2, cytotoxic T-lymphocyte-associated serine esterase 1)	210164_at	-3.241	0.047	yes	no	no
CX3CR1	chemokine (C-X3-C motif) receptor 1	205898_at	-3.105	0.047	yes	no	no
C17orf85	chromosome 17 open reading frame 85	236384_at	3.172	0.047	yes	no	no
RNF41	ring finger protein 41/E3 ubiquitin-protein ligase NRDP1	201962_s_at	3.379	0.047	yes	no	no
TNFSF13B=BAFF	tumor necrosis factor (ligand) superfamily, member 13b	223502_s_at	-3.365	0.047	yes	yes	no
CST3	cystatin C	201360_at	-3.257	0.048	yes	no	no
PBX1	pre-B-cell leukemia homeobox 1	212148_at	-3.294	0.048	no	no	no
TMEM170B	transmembrane protein 170B	235798_at	3.239	0.048	yes	no	no
ATP2B1	ATPase, Ca++ transporting, plasma membrane 1	209281_s_at	3.079	0.048	yes	no	no
RNF166	ring finger protein 166	227726_at	-3.140	0.048	yes	no	no
PPP2R2B	protein phosphatase 2, regulatory subunit B, beta	213849_s_at	-3.284	0.048	yes	no	no
IFT52	intraflagellar transport 52 homolog (Chlamydomonas)	233532_x_at	3.146	0.048	no	no	no
DIXDC1	DIX domain containing 1	214724_at	-3.303	0.048	no	no	no
IFT57	intraflagellar transport 57 homolog (Chlamydomonas)	222519_s_at	3.197	0.048	yes	no	no
C5AR1	complement component 5a receptor 1	220088_at	-3.054	0.048	yes	no	no
MARK4	MAP/microtubule affinity-regulating kinase 4	55065_at	3.272	0.049	yes	no	no
LOC202781	uncharacterized LOC202781	235587_at	3.255	0.049	no	no	no
WDSUB1	WD repeat, sterile alpha motif and U-box domain containing 1	226668_at	3.103	0.049	yes	no	no

TBC1D19	TBC1 domain family, member 19	220260_at	3.158	0.049	yes	no	no
GPR176	G protein-coupled receptor 176	227846_at	-3.219	0.049	yes	no	no
MVP	major vault protein	202180_s_at	-3.304	0.050	yes	no	no

Table S3: Gene coefficients and thresholds for score calculation using Affymetrix or Nanostring data.

This table presents the individual gene coefficients and the score threshold for measurements obtained from Affymetrix or recalibrated to the Affymetrix scale as presented in the article (left column) and an equivalent version of the algorithm with gene coefficients directly applicable to measurements obtained from Nanostring and associated to a corresponding threshold (right column).

Gene	Affymetrix coefficient	Nanostring coefficient
GADD45A	0.02342783269374480	0.02364299606154850
FOXO1	0.02238715554923010	0.01871616929277070
DCAF12	0.01992318764602770	0.01697620161900320
VPREB1	0.01785216833468160	0.01291751849231210
TAGAP	0.01777281264050780	0.01505935472245900
ORAI2	0.01767392311972970	0.01441065457090290
CXCR4	0.01716028348300170	0.01854984706253430
SEMA4B	0.01643084302272660	0.01100042323426500
AFF3	0.01436837353141670	0.01309012827710560
E2F5	0.01366650077542850	0.01216261363117030
KIAA0040	0.01289880819516210	0.01184366502462650
TCF4	0.01138291729747340	0.00968988480613538
USP44	0.00888287721788614	0.00680363017869830
PRDM15	0.00767507509953503	0.00837407330492499
FCRL2	0.00680499921125263	0.00807041962517729
RASSF6	0.00543954194264180	0.00705295029583474
EML6	0.00502834116159854	0.00465330524332114
SHISA8	-0.00885337137128934	-0.00917563069165676
ABCB1	-0.01422663288110540	-0.01516223771773760
VCL	-0.01464457146305320	-0.01726657013010330
ALDH2	-0.01674761157734610	-0.01587381890940880
METRNL	-0.02182703968248580	-0.02109080699425380
RGS10	-0.02707678406830700	-0.01913071555265900
Threshold	1.07507683691195000	1.37348744184531000

Table S4: Leading genes of ICA13 signature of the independent component analysis.

Genes are ordered by positive (1st part of the table) and negative (2nd part of the table) weight and by alphabetical order. The ICA13 positive leading genes overlapping with the PFS-supervised analysis are marked with * or ** (** more specifically showing genes that were retained in the final 23-gene predictor).

Genes with positive weight				**Genes with negative weight**
AACS	DKFZP761C1711	LOC100996301	SLAMF6	AMPD3
AICDA	DTX1	LOC283454	SLC25A27	CAP1
AMN	DUSP14	LRP2BP	SLC35E3	CTNNBL1
ANKRD36BP2*	E2F5**	LRRC37A3	SMAD1	CYTH1
APBB1	EIF2AK3	LTBP1	SPAG6	DDIT4
APBB2	EVA1B	MAD1L1	ST14	ECHS1
ARID1B	EVC2	MEGF11	ST6GALNAC4	GAPT
ARL15	FANCA	MID1IP1	SUSD3	GPR114
AVIL	FBXO15	MLK7-AS1	TAGAP**	INF2
BACH2	FBXO18*	MTSS1	TAPT1*	ITGB2-AS1
BCR	FHOD3	NCK2	TAPT1-AS1	MACC1
BIK	FRY	NFATC4	TBC1D1	PAOX
BMP3*	FST	NKX6-3	TCF4**	PARP15
BMP7	GABARAPL1	NSUN7	TCL6	PPDPF
C3orf52*	GNAZ*	ORAI2**	TDRD6	RASSF2
C4orf32	GOT2	P4HA2	TJP2	RPS18
C6orf99*	GPR137C	PEX5*	TMC5	THRB
C7orf10	GRSF1	PLA2G15	TMCC2	TLR10
C8orf12	GUCA1A	POU2AF1	TRIM55	TMEM109
CAB39L*	HDAC7	PRDM15**	TUBB2A	VASP
CCDC110	HERPUD1	PRKAR2B	USP44**	ZMIZ1
CDHR3*	HRK	RASAL1	USP53	
CDKN2C	IFT81	RASSF6**	VTCN1	
CHMP7	IGHG1	RBM38	WASF1	
CHPT1	IL4R	RCBTB1*	WDR19*	
COLEC11	INO80C	RGCC	WNT5A	
CPLX1	ISG20	RHOBTB2	XPR1	
CSNK1E*	KANK1	RIMKLB	YPEL5	
CXCR4**	KANK2	RIMS2	ZCCHC18	
CYP39A1	KCNH2	RIMS3	ZNF385B	
CYTH3*	KIAA0226L	RNFT2	ZNF804A	
DAB2IP	LDB3	RRM2B		
DCAF12**	LOC100130987	RTDR1		
DHTKD1*	LOC100506844	SIM1		

Table S5: Significant associations between ICA13 signature and gene sets from the LLMPP and MSigDB databases.

The "overlap" column shows: number of overlapping genes between the two sets/number of genes in signature/number of genes in reference gene set. *Abbreviations: FDR, false discovery rate; LLMPP, Leukemia Lymphoma Molecular Profiling Project.*

Source	Weight of leading genes	Category_Name	FDR q-value	overlap
LLMPP	positive	GCB-4 (GC_B_cell_BLhigh_DLBCLlow)	1.43E-10	10/133/35
LLMPP	positive	GCB-1	3.34E-06	16/133/306
LLMPP	positive	PanB-2	8.42E-05	8/133/82
LLMPP	positive	Tdiff-2	9.00E-04	6/133/56
LLMPP	positive	GCBDLBCL-3	9.85E-04	12/133/292
LLMPP	positive	CD40Dn-1	1.25E-03	6/133/64
LLMPP	positive	ABCDLBCL-1	1.25E-03	4/133/20
LLMPP	positive	MycCHIP-2	2.09E-03	43/133/2612
LLMPP	positive	PanB-3	2.63E-02	5/133/79
LLMPP	positive	TActDn-6	4.98E-02	2/133/7
MSigDB_C2	positive	HUMMEL_BURKITTS_LYMPHOMA_UP	9.54E-12	12/133/42
MSigDB_C2	positive	HADDAD_B_LYMPHOCYTE_PROGENITOR	1.47E-06	17/133/275
MSigDB_C2	positive	BASSO_CD40_SIGNALING_DN	3.73E-03	7/133/66

Table S6: Nucleotide sequences of probes used for NanoString experiments, for the 23 genes retained in the final predictor and 4 housekeeping genes.

Gene	Ensembl Genome database release 86 oct 2016	Sequence_Probe A NANOSTRING	Sequence_Probe B NANOSTRING
ABCB1	ENSG00000085563.14	CTAACAAGGGCACGAGCTATGGCAATGCGTTGTTTCTGGCCACCAGAGAG CCGTCTCAGATGAGTGGGTTAATCAATCAAGTATG	CGAAAGCCATGACCTCCGATCACTCCTGTATCCAGAGCTGACGTGG CTTCATCCAAAAGCAAAATATGAGGCTGT
AFF3	ENSG00000144218.18	TTCAGGCTCCTACTGAAGGGACAATCTAGGTAATGGCCTCATTTAGCCAAC TTTCGGGTTATATCTATCATTTACTTGACACCCT	CGAAAGCCATGACCTCCGATCACTCGAGAATCAGCATTATCATGG CAAATAGCTCTCAGCCTGGTAAGCGAAGGA
ALDH2	ENSG00000111275.12	TGCTGATCTTGCTGAACTTTCCATCAATGGCTGAGGGAGGAAGCTTGCATC AACAGCCACTTTTTTTCCAAATTTTGCAAGAGCC	CGAAAGCCATGACCTCCGATCACTCTTTCTCTTTTCAGATATTCAG CACGCAAGGATCATTTTTCTTGGTTTTGT
CXCR4	ENSG00000121966.6	ATGAAGGAGTCGATGCTGATCCCAATGTAGTAAGGCAGCCAACAGGCGAA CACAAGAATCCCTGCTAGCTGAAGGAGGGTCAAAC	CGAAAGCCATGACCTCCGATCACTCTGTGCACAGTGTTCTCAAACT CACACCCTTGCTTGATGATTTCCAGGAGG
DCAF12	ENSG00000198876.12	AATTCCAGAACCTGTGGTATGTGCTGTGTAACAAGTTAGGGTGGACTTGCC ATCTCCATGACTGCTTGAGCGGCTGGAGAATCTG	CGAAAGCCATGACCTCCGATCACTCTTGCTTAATTCCCTGTTTATA GCACCTATTTCTCTAGGTTCGAAGATGAG
E2F5	ENSG00000133740.10	GGAGTCACTGGAGTCAAGGACTGGGAGGAAGGCTGTGTGAGCTTGGAGGA GTTGATAGTGGTAAAACAACATTAGC	CGAAAGCCATGACCTCCGATCACTCATGTTGCTCAGGCAGATTTTG AGTTGCCATGCTGGATTTCTGT
EML6	ENSG00000214595.11	TTCCCATAAGCAGGTAACCGATGTCTTTCATTTAGGAAGCCTTGGCGTGAC AGCAAGAAGGAGTATGGAACTTATAGCAAGAGAG	CGAAAGCCATGACCTCCGATCACTCAAACCCATAATTCTACCCTGT CCCGAACTTCAATTCCTTTGCTGCTCACC
FCRL2	ENSG00000132704.15	GAAACCAAGGACACCAAACAGTCCCCAGAGAACTCCAGCTGTCATGAGGT CGCTATGCAGACGAGCTGGCAGAGGAGAGAAATCA	CGAAAGCCATGACCTCCGATCACTCTCTCCTGATATCTTGTGGAAC AAGGCATACAACAGCAAAGCAACACCAGT
FOXO1	ENSG00000150907.7	GGTGCCAGGTGAGGACTGGGTCGAAACAGTTAATGATGTTGGTGATGAGA CACCAGTTAGCGTGGCGTATACCATGTTGTTAACA	CGAAAGCCATGACCTCCGATCACTCAAACTGGTGTTTGGTGGCGCA AACGAGTAGCACGGCGTCTGCTGCATCAT
GADD45A	ENSG00000116717.11	TTGAACTCACTCAGCCCCTTGGCATCAGTTTCTGTAATCCCTGAATCAATAG AACAATATCAGTTATGGCGGTG	CGAAAGCCATGACCTCCGATCACTCCATCTGCAAAGTCATCTATCT CCGGGCCCCCAGAACATGTAG
KIAA0040	ENSG00000235750 .9	AAGGGTCACCACTAAATGCCCACCCTCCACAGGCCTAAAGTACACCAAAT GCACTCTATATGGAGGGAGAGTAGCTGGAT	CGAAAGCCATGACCTCCGATCACTCGCAGGTTTAAATTATGATTCA CAGGGTCATCTGTTAGAAAACCCTGGTGC
METRNL	ENSG00000176845.12	ATAGAGTCTGCTCACGCGCAGGTGGATGGCTGAGTCATGTTGGAGTTAACG GAGACCCGCCATCGTTTAC	CGAAAGCCATGACCTCCGATCACTCTCGGGCACCGGCTCGAAGAC CCTGCTTTTCTGCCG
ORAI2	ENSG00000160991.15	TACATAACGCCATGCATAGATCATCAGCACGAGGCAGGTAAGGGCTGCAG CTATCAGCTAATAGGGTCGGCTCAACAGTGTATCC	CGAAAGCCATGACCTCCGATCACTCGTTTTAAAAACCCGCACACA GGACAGTGGCTGTCTCTGCCGCACGTGATC
PRDM15	ENSG00000141956.13	TGACCCCTTGCACATGATGACCGTGATGACCTCGAAGGTTGCAGGATCAGC AGATAAGGTTGTTATTGTGGAGGATGTTACTACA	CGAAAGCCATGACCTCCGATCACTCTTCATGGATATTACTAAGATG CGTTTTGAGCTGGAGTCAGTGACCTCACG
RASSF6	ENSG00000169435.13	CCTTCAACCAGTTTAGGTTTGGTAATACTGAGAAGTAATGCCTTTGTCTTCT AGCCCAGATCCTACGAGATGAGCTACGTAACTA	CGAAAGCCATGACCTCCGATCACTCACTCAGTAAGAGAGACTTCTT AGGCTTGAAAAGGCCATAGCACACTTTCA

RGS10	ENSG00000148908.14	GCCTTGCTGGACAGAAAGGTCATGTAGATCTCCTTTGCCTTTTCCTCAGGTTGTTACTTGAAGGGTTCAACACGAGCTC	CGAAAGCCATGACCTCCGATCACTCTGAGCCGAGACTGCCCCTCCACGTTGACCTGTGATGAG
SEMA4B	ENSG00000185033.14	GAGGGAAAGGCGTACATTGTCACGAAATCTCTCCTGAGCTCGCGCTGCCTGCATTCTCATGGAAATGCAATGGATTCATTCC	CGAAAGCCATGACCTCCGATCACTCTCACGCAACAACGGAGGAAGGCAGGCGACAGTCTCTTCCCTGAATTCT
SHISA8	ENSG00000234965.2	TAGGTACCGGGACCCGCGGCCCGCGCTGCCGTAAACCTGTTGCAGTATCACGTAAATACCTACTTCGATA	CGAAAGCCATGACCTCCGATCACTCCTTCACACGGTGACCTCGGTCTTGCTATTGGTCCT
TAGAP	ENSG00000164691.16	TGGAATTGGTCAGCCTCAAAGACCGGCTGGCTACATCGTCGCACGAGACACCTTCTGGAATTTCTTCCTTTGATTTTGCCATTTT	CGAAAGCCATGACCTCCGATCACTCTATGGCATGGCGTATGGCCTCCCTCCTAAATATACGATTCTTTGGCATAT
TCF4	ENSG00000196628.14	GTCGCTGTGCCATAGCTTGATCTATGTTTCTATCAAATATGGTAAGTATACGTCGTGTCTTAGACGACTGTGTGTGATTCTCGAG	CGAAAGCCATGACCTCCGATCACTCATGGAAGTTGGTTCAATTGTGCCCAGACGTCAAAGCTAGCTATCTGATGA
USP44	ENSG00000136014.11	CTCCCCAACAGGAGCTCTGGAGGCAAAAGTTTAGAATGTCCATTCTCAGTCCTTCTTGAAGACCTATGTAAAGAAACGGGTCACT	CGAAAGCCATGACCTCCGATCACTCAGCTAAGGATTTCATTAGACGAGGTATCAGCGTCTTCATTGGGATGTTGG
VCL	ENSG00000035403.16	AATATGAAACCCATTACTCTCTTGAACCCCATACTTGGGAGTAGATGCAGCCATACGCATGACTACATTACAACGGGCCAGGAAG	CGAAAGCCATGACCTCCGATCACTCAAACACTGCCCTCTCATTTTGCCTAGTAGGAACTTACTGTGGTGATAAGA
VPREB1	ENSG00000169575.4	CAGGTGAGGCGGATTGTGGTTCCAAGGGCCGAGGACATGCAGGTCTCGATCTCGTACAAAACGACTATGACCAT	CGAAAGCCATGACCTCCGATCACTCTAGACGCTGTACACACCGATGTCATGGTCGTTCCTCAGGGTG
HOUSEKEEPING GENES			
ACTB	ENSG00000075624	GATCTTGATCTTCATTGTGCTGGGTGCCAGGGCAGTGATCTCCTTCTGCACTGTTGAGATTATTGAGCTTCATCATGACCAGAAG	CGAAAGCCATGACCTCCGATCACTCAGGATGGAGCCGCCGATCCACACGGAGTACTTGCGCTCAGGAGGAGCAAT
GAPDH	ENSG00000111640	AAGTGGTCGTTGAGGGCAATGCCAGCCCCAGCGTCAAAGCCTCAAGACCTAAGCGACAGCGTGACCTTGTTTCA	CGAAAGCCATGACCTCCGATCACTCCCCTGTTGCTGTAGCCAAATTCGTTGTCATACCAGGAAATGAGCTTGACA
MRPS9	ENSG00000135972	TGCTTCATTACTGGCCTGGCTCGTTTCTCAAACAAACCACTTGGGAAAAGCAAAGACGCCTATCTTCCAGTTTGATCGGGAAACT	CGAAAGCCATGACCTCCGATCACTCCATCTTCTCCCCACTGGATTGCTCTTTGTCTTGGAAAAATCTGTTCAGGA
PGK1	ENSG00000102144	ATTCCACACAATCTGCTTAGCCCGAGTGACAGCCTCAGCATACTTCTTGCCACAATTCTGCGGGTTAGCAGGAAGGTTAGGGAAC	CGAAAGCCATGACCTCCGATCACTCGCTTTGGTTCCCCGGGCAAAAGCTTCCCATTCAAATACCCCCACAGGACC

References

1 Salles G, Seymour JF, Offner F, *et al.* Rituximab maintenance for 2 years in patients with high tumour burden follicular lymphoma responding to rituximab plus chemotherapy (PRIMA): a phase 3, randomised controlled trial. *Lancet* 2011; **377**: 42–51.

2 Huet S, Xerri L, Tesson B, *et al.* EZH2 alterations in follicular lymphoma: biological and clinical correlations. *Blood Cancer J* 2017; **7**: e555.

3 Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010; **33**: 1–22.

4 Hothorn T, Lausen B. On maximally selected rank statistics. R News. 2002; : 3–5.

5 Rotival M, Zeller T, Wild PS, *et al.* Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. *PLoS Genet* 2011; **7**: e1002367.

6 Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. *Neural Netw Off J Int Neural Netw Soc* 2000; **13**: 411–30.

7 Liebermeister W. Linear modes of gene expression determined by independent component analysis. *Bioinforma Oxf Engl* 2002; **18**: 51–60.

8 Shaffer AL, Wright G, Yang L, *et al.* A library of gene expression signatures to illuminate normal and pathological lymphoid biology. *Immunol Rev* 2006; **210**: 67–85.

9 Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R. Stat Soc Ser B. 1995; : 289–300.

10 Dave SS, Wright G, Tan B, *et al.* Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells. *N Engl J Med* 2004; **351**: 2159–69.

11 Casulo C, Byrtek M, Dawson KL, *et al.* Early Relapse of Follicular Lymphoma After Rituximab Plus Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone Defines Patients at High Risk for Death: An Analysis From the National LymphoCare Study. *J Clin Oncol Off J Am Soc Clin Oncol* 2015; **33**: 2516–22.

12 Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nat Genet* 2005; **37**: 382–90.

13 Hystad ME, Myklebust JH, Bø TH, *et al.* Characterization of early stages of human B cell development by gene expression profiling. *J Immunol Baltim Md 1950* 2007; **179**: 3662–71.

14 Dybkær K, Bøgsted M, Falgreen S, *et al.* Diffuse large B-cell lymphoma classification system that associates normal B-cell subset phenotypes with prognosis. *J Clin Oncol Off J Am Soc Clin Oncol* 2015; **33**: 1379–88.

15 Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinforma Oxf Engl* 2003; **19**: 185–93.

16 Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; **56**: 337–44.

17 Pangault C, Amé-Thomas P, Ruminy P, *et al.* Follicular lymphoma cell niche: identification of a preeminent IL-4-dependent T(FH)-B cell axis. *Leukemia* 2010; **24**: 2080–9.